CRWR Online Report 98-7

Development of a GIS Database for Estimating Biogenic Hydrocarbon Emissions in North Central Texas

By

Irvin Wade Strange, M.S.E.,

David T. Allen, PhD., and David R. Maidment, PhD.

May 1998

CENTER FOR RESEARCH IN WATER RESOURCES

Bureau of Engineering Research • The University of Texas at Austin
J.J. Pickle Research Campus • Austin, TX 78712-4497

This document is available online via World Wide Web at
http://www.ce.utexas.edu/centers/crwr/reports/online.html

ACKNOWLEDGEMENTS

I would like to thank Dr. Allen for giving me the opportunity to conduct this research under his supervision. He is inspiring in his pursuit of higher air quality, and he has made the work interesting and exciting. I would also like to thank Dr. Maidment for his continual guidance in the ever-evolving field of geographic information systems. I consider myself very lucky to have had the opportunity to study under two men of such great influence in their respective fields. Dr. Kinney also deserves my gratitude for taking me under her wing for a short time and for her guidance on this thesis.

I would like to thank all of my fellow students who have helped me over the course of my studies. Graduate students working for both Dr. Allen and Dr. Maidment were invaluable with their unselfish donations of time and assistance. I would especially like to thank Christine Wiedinmyer for all of her effort and enthusiasm in this work.

Lastly, I would like to thank my family for all the support they have given me. My mother Nancy, my father Glenn, and my sister Monica have always appreciated my efforts and encouraged me to do my best. None of my academic accomplishments would be possible without the support of such a great family.

May 6, 1998

Development of a GIS Database for Estimating Biogenic Hydrocarbon

Emissions in North Central Texas

by

Irvin Wade Strange, M.S.E.

The University of Texas at Austin, 1998

Supervisors: David T. Allen and David R. Maidment

Biogenic hydrocarbons are a significant fraction of the total VOC emission inventory for the United States and contribute to the formation of tropospheric ozone. However, the amount of the biogenic hydrocarbons being emitted into the atmosphere is not precisely known. The current biogenic emission estimation models and urban airshed models use low-resolution data for estimating vegetation biomass and VOC emissions.

During this research, the geographical information systems ArcInfo and Arcview were used to construct a high resolution, spatially accurate vegetation biomass and biogenic emissions database for the Dallas/ Fort Worth area. Many sources of data on vegetation cover were evaluated, and the final mappings incorporated information from three major data sources.

The methodologies developed in this work are presented along with the final composite database for 37 counties in North Central Texas. Preliminary emission plots for the domain were developed and are presented. The strong dependence of biogenic emissions on the spatial distribution of vegetation species made the use of a GIS an ideal method for estimating biogenic emissions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

High concentrations of the air pollutant ozone (O_3) are widespread in the troposphere, and reducing ozone concentrations remains one of the most difficult air quality challenges facing the United States. Limiting the emissions of the reactants that form ozone in the troposphere can reduce ozone levels. Oxides of nitrogen (NOx) and volatile organic compounds (VOCs) are the chemicals that react in the presence of light to form ozone.

One of the major sources of VOC emissions which has been historically underestimated is biological volatile organic compounds (BVOCs). On a global scale, BVOCs contribute more to VOC emissions than do all anthropogenic hydrocarbons. Because of their high reactivities BVOCs may play a significant, if not dominant, role in the formation of ozone and other air pollutants for many parts of North America.

1.1 BIOGENIC EMISSION FACTORS

The following factors affect the quantity and type of hydrocarbons emitted by vegetative sources: light intensity; temperature; plant species; biological stress; water stress; and the amount of plant biomass. The effect of all of these factors on the production of biogenic hydrocarbons (BVOCs) has been studied to some extent, as described in the following section.

The main factors that affect the amount of isoprene or other BVOC that a plant emits are temperature, light, and plant species. Guenther *et al.* (1993) investigated the sensitivity of several available isoprene and monoterpene emission rate models to temperature, and they suggested a simple exponential relationship between leaf temperature and monoterpene emissions. Guenther *et al.* (1993) observed increases in isoprene emissions depending on light intensity and temperature with a maximum emission rate at about 40 ^{o} C.

The intensity of light falling on the foliage of an emitting tree is difficult to determine. This variable is sometimes addressed through the use of canopy models or scaling relationships to account for the changes in solar radiation as a function of height in a tree's canopy (Lamb *et al*, 1993).

The rate and type of BVOC emission is highly dependent on the species of the emitting plant. Several researchers have published BVOC emission rates for specific plant species. The most common method used to obtain these rates is to enclose an emitting portion of the plant, a technique that is sometimes referred to as "bagging a tree." The leaves, branches, or the entire tree is enclosed to make a sealed control volume, and the BVOC emissions from the plant are then monitored while keeping the internal conditions constant. The difficulty with these experiments is that microclimate changes can affect the emission of the chemicals. Fuentes *et al.* (1996) used environmentally regulated leaf cuvettes to collect emission samples from aspen trees. Extensive emission rate studies have been conducted on vegetation in California's South Coast Air Basin (Arey *et al.* 1995, Winer *et al.* 1992). Other emission rate data have been published for various vegetation species from around the world. Benjamin *et al*. list several articles by various researchers that contain published emission rates (Benjamin *et al*. 1996). Other researchers have developed emission rate estimation methods based on the genus and species of plants. If emission rates were not published for a particular plant of interest, emission rates were assigned based on the emission rates of plants with similar species. Benjamin *et al.* (1996) have assigned emission rates to 377 southern California species by using a taxonomic method to develop emission rates for 253 plants from 124 measured species. Table 1.1 shows some of the typical measured and published isoprene and monoterpene emission rates used by Benjamin *et al.*

Table 1.1: Sample vegetation species with published emission rates (Benjamin *et al*. 1996)

Botanical Name	**Common Name**	**Isoprene Rate (μg(g leaf)$^{-1}$h^{-1})**	**Monoterpene Rate (μg(g leaf)$^{-1}$h^{-1})**	**Reference**
Prunus avium	Bing Cherry	NED	0.1	Winer *et al.* 1992
Pinus ellotii	Slash pine	NED	5.0	Tingey *et al.* 1980
Quercus dumosa	California Scrub Oak	5.2	0.0	Arey *et al*. 1995
Quercus wislizenii	Interior Live Oak	12.5	0.0	Arey *et al*. 1995

(NED = No emissions detected)

Correct isoprene and monoterpene emission rates are essential for the study of BVOCs. In addition to the obvious contribution to accurate emissions inventories, some researchers suggest that these emission rates should be taken into account for urban planning so as not to add unnecessary emissions to urban

VOC inventories (Benjamin *et al.*, 1998). Benjamin *et al.* (1998) have quantified the ozone producing potential of many Southern California tree species by taking into account the hydrocarbon emission rates, the typical foliar biomass for each plant species, and the maximum incremental reactivities for isoprene and monoterpenes. Tree types with low ozone forming potential (OFP) are identified and recommended for large planting operations in cities with poor air quality.

One major uncertainty associated with using emission rates found in the literature is that the rates must be corrected for temperature and light differences depending on the conditions for which the rates are to be applied and the conditions under which they were established. There are atmospheric correction correlations available to account for these differences. The most widely used correction algorithms are G93 published by Guenther (Guenther *et al.*, 1993) and T400 and T800 published by Tingey (Tingey *et al*., 1980).

Some researchers, however, feel that the algorithms that only take into account the temperature, light exposure, and species of emitting vegetation are not sufficient for VOC emission estimation. Monson *et al.* (1995) argues that there are still many variables that need to be considered. Monson states that a connection to plant physiological mechanisms such as water, carbon, and nutrient cycling is needed to understand the true relationship between the atmosphere and the biosphere. Monson also warns against applying emission rates that were obtained under ideal, ‘greenhouse’ conditions to vegetation under non-ideal conditions.

A BVOC emission rate is expressed in units of grams of chemical produced per gram of biomass per unit of time. Therefore, the amount of foliage or leaf biomass of the emitting plant determines how much BVOC is produced. Several researchers including Winer at UCLA, Nowak, and Geron at Research Triangle Park have published algorithms to simplify the calculation of a plant's foliage. Geron developed equations to estimate leaf area and leaf biomass based on a tree's crown parameters and on a tree's diameter at breast height (dbh). The results of the equation based on crown parameters (diameter, height, and outer surface area) were found to be more representative of the trees investigated (Geron *et al.,* 1994). For a Texas Natural Resource Conservation Commission biogenic estimation project conducted in Houston, tree biomass was estimated using a tree's crown characteristics. Researchers used biomass algorithms which required the estimation of a radially symmetrical crown and the application of a shape and clump factor (Estes *et al*., 1995).

1.2 GIS AND BIOGENICS

Biogenic hydrocarbon emissions are highly dependent on temperature, species, and light distributions and are therefore highly dependent on location. From region to region, many parameters fluctuate, thus changing the species and make-up of the natural vegetation. The recent trend in biogenic volatile organic compound modeling is to utilize a spatial database for storing emission factors, calculating emissions estimations, and running photochemical models. This trend is exemplified by studies done by Benjamin *et al*. (1997) and Kinnee *et al.* (1997).

A Geographic Information System is one such spatial database that is beneficial in all of these tasks.

Some researchers are beginning to develop their own systems for using GIS with BVOC emissions estimations. Benjamin *et al*. have constructed an extensive database in a GIS for California's South Coast Air Basin (SoCAB). As of 1997, the database consisted of spatial and temporal data to help predict the VOC emissions of 287 species of vegetation found in the domain.

Benjamin *et al*. (1997) made several improvements upon previous biogenic emission studies done in California's SoCAB. These improvements include a scale of 50m and 333m for the urban and rural land use regions respectively, recently measured BVOC emission rates, improved environmental correction algorithms, and an empirically based lapse rate to adjust temperature with elevation (Benjamin *et al*., 1997). Much of the database created by Benjamin *et al.* was developed by digitizing paper maps and spreadsheets and converting the information into a grid cell array for a GIS software called IDRISI.

Another recent biogenics study that incorporated a Geographic Information System for a much larger area than the SoCAB study was conducted by Kinnee *et al*. (1997). In their study, Kinnee *et al* (1997) built a GIS database for the entire continental United States using electronically available land use/ land cover vegetation data. As a result of this effort, they developed the Biogenic Emissions Landcover Database (BELD) to describe the vegetation of the eastern United States. The BELD was constructed on a county scale from nine sources. The source used for a particular area depended on a hierarchy of rules established by the research team. Some of the dominant sources included the U.S. Forest Service Eastwide database, the USGS Land Cover Characteristics database (inland water and vegetation classes), the U. S. Census Bureau's 1990 urbanized area boundaries, and the U. S. Department of Commerce 1987 Census of Agriculture. Less data were available for the western half of the United States; therefore, the United States Geological Survey's (USGS) Land Cover Characteristics (LCC) data set was used almost exclusively in that region (Kinnee *et. al*, 1997). These data are based on satellite data and are available in 1-km grid cells for the entire continent. The USGS LCC data set is discussed in more detail in the methodology of this document.

There are many possible applications for GIS which would be beneficial to any emissions estimating efforts. Without even considering the remote sensing technologies which are compatible with GIS, there are many parameters that affect BVOC emissions that can be stored and shown spatially using a GIS. As stated before, the three main factors that BVOC emissions depend on are temperature, light intensity, and vegetation species. Parameters that would affect the temperature of a plant and can be incorporated into a GIS database include ground elevation, diurnal temperature changes, seasonal variation, and topography (e.g., the sunny or shaded part of a mountain). Some parameters that would affect the light intensity of a plant include canopy shading, topography, and seasonal variation. Soil type, precipitation, water availability, elevation, and vegetation cover are all parameters that affect the type of species found in any one area. The main focus of the current research effort is to identify the existing vegetation species distribution as accurately as possible.

1.3 STUDY OVERVIEW

The Texas Natural Resource Conservation Commission (TNRCC) is updating its emission inventory of volatile organic compounds (VOCs) for the Dallas-Fort Worth ozone nonattainment area. Biogenic hydrocarbon emissions are known to be important contributors to VOC emissions in the North-Central Texas area, but the emission rates are highly uncertain. The uncertainty is due to a number of factors including unknown distribution of vegetation types. In response to the concerns about uncertainties in biogenic emissions, TNRCC contracted with ENVIRON International Corporation, the National Center for Atmospheric Research (NCAR), and the University of Texas at Austin to prepare a biogenic emissions estimation database for a 37 county North Central Texas domain in a format compatible with photochemical and emission modeling requirements. The modeling domain is shown in Figure 1.1. The counties in the domain are listed in Table 1.2. The high temperatures and changing wind directions make the BVOCs being emitted from the vegetation surrounding Dallas and Tarrant counties very important in understanding air pollution formation in the Dallas/ Fort Worth Metroplex. In particular, the large stands of deciduous and pine forest to the south and east of Dallas county are important because of the potential for wind transport of large emissions from this area to the Dallas/Fort

Worth airshed. Here the BVOCs will have the opportunity to react with high levels of NOx in the urban atmosphere to form O_3.

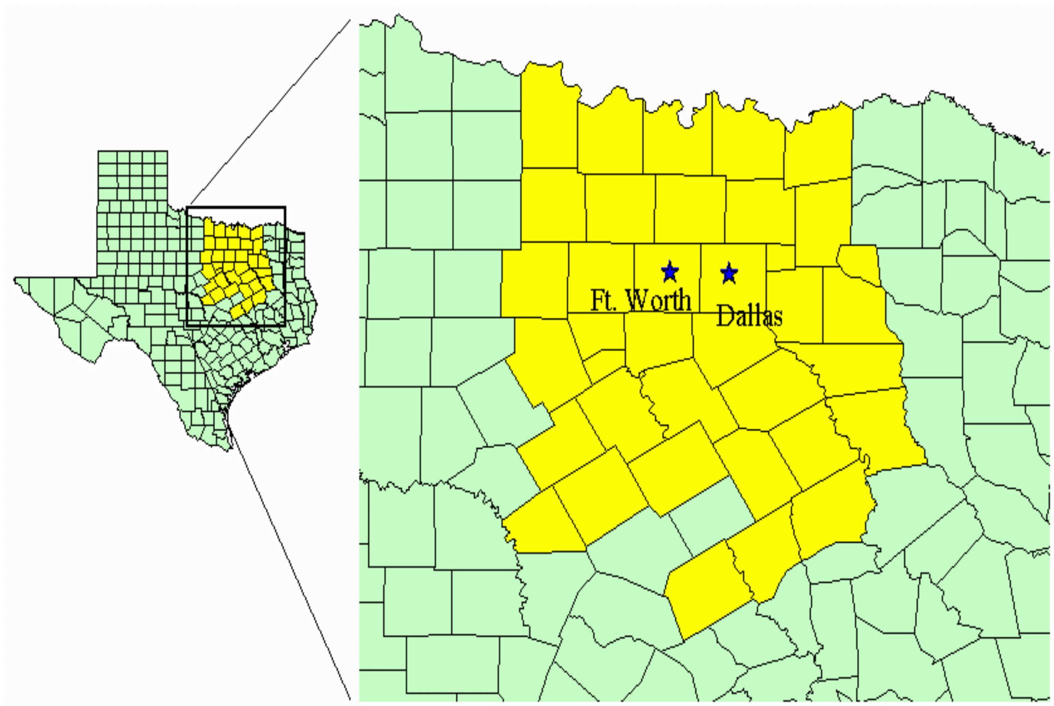


Figure 1.1: 37 County Biogenic Emissions Study Domain

Table 1.2: 37 counties in North Central Texas study domain

COUNTY	**FIPS County Code**	**COUNTY**	**FIPS County Code**
Anderson	1	Johnson	251
Bosque	35	Kaufman	257
Clay	77	Lampasas	281
Collin	85	Leon	289
Cooke	97	Limestone	293
Coryell	99	McLennan	309
Dallas	113	Milam	331
Denton	121	Montague	337
Ellis	139	Navarro	349
Erath	143	Palo Pinto	363
Fannin	147	Parker	367
Freestone	161	Rains	379
Grayson	181	Robertson	395
Hamilton	193	Rockwall	397
Henderson	213	Somervell	425
Hill	217	Tarrant	439
Hood	221	Van Zandt	467
Hunt	231	Wise	497
Jack	237		

The first task for this study was to develop a vegetation database. The goal was to create a spatial database with appropriate land use or land cover information so that distribution of vegetation species could be represented. It was decided that a GIS would be used to construct the database.

1.4 THESIS OBJECTIVES

This thesis discusses the methodologies and rationale used in the development of a North Central Texas vegetation database in a GIS format. The validity and applications of the final database are also discussed. The purpose of the database is to serve as a basis for estimating biogenic hydrocarbon emissions from the 37 county North-Central Texas study domain. The geographic information systems used for this development were Arc/Info and Arcview. The following chapters will discuss the applications, advantages, and future uses of this GIS vegetation database for estimating the amount of biogenic hydrocarbons released into the atmosphere, particularly for the North-Central Texas study domain.

2. METHODOLOGIES

2.1 SOURCES OF DATA

GIS compatible Land Use and Land Cover (LULC) data are available for the North-Central Texas region from the North Central Texas Council of Governments (NCTCOG) Land Use/Land Cover database, a Texas Parks and Wildlife database, the United States Geological Survey's (USGS) LULC data files and several other sources. Based on differences in the types and quality of data available for rural and urban regions, it was decided that the urban and rural vegetation distributions would be considered separately. The rural GIS land cover coverages were designed to differentiate between the separate vegetation types for the rural areas. However, the urban land use coverages gave little if any information on the urban vegetation. The strategy for the urban areas, especially for the Dallas/Ft. Worth metropolitan area, was to use the land use classifications for the database development and to add field survey information about the vegetation found in that area. The third type of electronic information used in the study was the county crop distribution information from the USDA. This information was not originally in GIS format but could be used to describe the agricultural crop distribution within each county in the study domain. The major sources of data considered for use in this study are summarized in Table 2.1 and briefly described below.

Table 2.1: Potential biogenic emission data sources.

DATA SET	SOURCE	Scale	Last Update
USGS Land Use Land Cover	USGS http://edcwww.cr.usgs.gov/Webglis/glisbin/searchchoiceftp.pl?dataset_name=1_250_LULC	1:250,000	1974-1978
Texas Parks and Wildlife Department Vegetation Coverage	TP&WD http://www.tnris.state.tx.us/gispage.html	1:250,000	1984
USDA-National Agricultural Statistics Service (NASS)	USDA http://usda.mannlib.cornell.edu/	County	1995, 1996
North Central Texas Council of Governments Land Use/Land Cover	NCTCOG	Derived from 1:250,000	1990
USGS Land Cover Characteristics (LCC)	USGS http://edcwww.cr.usgs.gov/landdaac/glcc/na_int.html	1 km	1993
Landsat Thematic Mapper Data	EROS or EOSAT http://edcwww.cr.usgs.gov/webglis	30 meter	Ongoing
BEIS-2 Forest Inventory Analysis database	US EPA http://www.epa.gov/asmdnerl/biogen.html	1/6 degree longitude by 1/9 degree latitude	July 1995
Texas Gap Analysis Database	Texas Cooperative Fish and Wildlife Research Unit http://www.tcru.ttu.edu/txgap/home/index.html	Unknown	In progress

The United States Geological Survey LULC files are standard land use/land cover assignment available for the entire United States. The data are reported in an Anderson level two classification system and are distributed in one degree latitude by two degree longitude tiles. The Anderson classification level refers to the number of times a LULC category is subdivided. A level two classification system has the Urban category subdivided into Residential, Commercial, and Industrial designations. A level three classification system has further subdivisions for each of the Residential, Commercial, and Industrial categories. The data can be found in electronic Arc/Info export format at http://edcwww.cr.usgs.gov/Webglis/glisbin/searchchoiceftp.pl?dataset_name=1_250_LULC of the world wide web. Another source for this data is the EPA ftp site located at ftp://ftp.epa.gov/pub/spdata/EPAGIRAS/.

Data in the form of a vegetation coverage for the entire State of Texas are available from the Texas Parks and Wildlife Department database. The data are available in Arc/Info export format from the TNRIS web site (http://www.tnris.state.tx.us/ftparea.html). The data have low spatial resolution relative to the USGS data; however, ground surveys described in this report indicate that these data have the most accurate vegetation classifications of the study region. The coverage was last updated in February of 1984, and it was originally constructed from Landsat imagery taken from 1976 to 1980.

The USDA-National Agricultural Statistics Service (NASS) provides data on crop harvests by county. The National Agricultural Statistics Service home page can be found at http://www.usda.gov/nass/nasshome.htm. Agricultural data are available for most crop producing counties in the United States.

The North Central Texas Council of Governments (NCTCOG) land use database was given to the project team in electronic format by the NCTCOG via the TNRCC. The database was derived from the USGS LULC data and was updated with 1990 LULC information for the area. This database extends the Anderson classification coding system from the standard level two system used by the USGS LULC to either a level three or level four system.

The USGS North America Land Cover Characteristics (LCC) database is part of a global vegetation database being prepared by the USGS. These data are on a 1 km by 1 km grid for the entire continent of North America and are presented in four different classification systems in two coordinate system projections. The most detailed classification system has 205 vegetation categories for the entire continent, of which approximately 50 are found within the North-Central Texas study region. The data are available in monthly composites spanning from April 1992 to March 1993 and 10 day composites for a slightly larger time frame. Each classification system is available in both Lambert Azimuthal Equal Area or Interrupted Goode Homolosine coordinate projections. These data were derived from Advanced Very High Resolution Radiometer (AVHRR) data by interpretation of a Normalized Difference Vegetation Index (NDVI). AVHRR data are high-resolution data produced by orbiting satellites. The NDVI composites depict the relative density or vigor of different vegetation types. The NDVI is calculated by taking the difference of near-infrared (AVHRR Channel-2) and visible (AVHRR Channel 1) reflectance values divided by total reflectance. These values are then scaled from 0 to 200 where values less than 100 are usually clouds, snow, or water. The NDVI is then composited with the

AVHRR data which involves creating a multi-band image with five AVHRR channels, the NDVI value, four bands of computed satellite/solar geometry, and a band which identifies the specific daily overpass from which each pixel was extracted. The data can be obtained from the ftp site at ftp://edcftp.cr.usgs.gov.

Landsat Thematic Mapper Data is produced from satellite imagery and is available commercially from either the EROS Data Center (605-594-6151) or EOSAT (1-800-232-9037). The EROS Data Center's images are mostly over 10 years old, while EOSAT has more recent data. The cost for one image is approximately $4400 for non-federal government project use. The USGS web site located at http://edcwww.cr.usgs.gov/webglis allows searches of the imagery. The high-resolution data can be used to obtain information regarding plant community distributions and relative density distribution ratios. This information can then be used for sampling site selection and QA/QC when evaluating other LULC sources.

The BEIS-2 Forest Inventory Analysis database was used in the development of the US EPA's Biogenic Emission Inventory System, Version 2 (BEIS-2). Information can be found at a website maintained by the US EPA: http://www.epa.gov/asmdnerl/biogen.html. The database contains measurements from 97,000 1-acre plot locations in the eastern US.

The Texas Gap Analysis database will contain a vegetation coverage of the state of Texas. The vegetation coverage is currently being built by ground-truthing Landsat Thematic Mapper data and classifying the vegetation for the state of Texas. This project is being done by the Texas Cooperative Fish and

Wildlife Research Unit. The database is incomplete at this time. Information on this data source can be found at http://www.tcru.ttu.edu/txgap/home/index.html/.

2.2 SOURCES OF LAND USE LAND COVER DATA NOT USED TO ESTIMATE LEAF BIOMASS DENSITY

Some of the databases were immediately determined not to be useful for the biogenic emission database development. The BEIS-2 Forest Inventory Analysis database was not used. Although there are default vegetation data that could be used for our study domain, the data are spatially coarse. The data used in developing this database did not include any samples from the North Central Texas domain. There may be some benefit in comparing the final results of this project to the BEIS default data, but the BEIS data were not used directly.

The Texas Gap Analysis database was not used in the development of this database as it is incomplete at this time. The projected completion date for the Texas Gap Analysis vegetation map is June 1998. Although it has potential to be a good source of information on the vegetation types of Texas, no data are currently available for the study region to be evaluated in this research.

The other six data sets identified in Table 2-1 as potential sources of vegetation data for the GIS database required some GIS processing before their value could be assessed. The development of all of these data sources will be discussed except for the Landsat Data as it was processed by investigators at the National Center for Atmospheric Research. The application of the data to the biogenic emission inventory is discussed in the results section of this document.

2.3 SOURCES OF LAND USE AND LAND COVER DATA USED TO ESTIMATE LEAF BIOMASS DENSITY

The following were the primary sources of data used in this study:

1) *The Texas Parks and Wildlife data on vegetation cover*

2) *The North Central Texas Council of Governments Land Use data*

3) *The USDA-National Agricultural Statistics Service (NASS) data*

The methods and rationales used in processing these data sources is described in more detail below. The database descriptions also include some discussion of the reasons for not using some of the other available information, such as the LCC database from the USGS.

North Central Texas Council of Governments Land Use data

The North Central Texas Council of Government's land use/land coverage database was developed in 1990 as an update and extension of the USGS LULC database. It covers the Dallas/Fort Worth area including Dallas county, Tarrant county, and parts of Wise, Jack, Denton, Collin, Rockwell, Kaufman, Ellis, Johnson, Hood, and Parker counties. The LULC categories for this database are shown in Table 2.2. The database was given to the project team by the NCTCOG via the TNRCC. The coverage was originally in a Universal Transverse Mercator (UTM) zone 14 projection and was converted to a UTM zone 15 projection. Although the domain is in zone 14, zone 15 coordinates were used for compatibility with previous biogenic emissions work done in the Houston area. The projection of a coverage refers to the system used to represent the surface of interest on a two dimensional plane. Each projection type has advantages in its method of converting the normally curved surface of the globe to a flat representation. The Universal Transverse Mercator (UTM) projection can be thought of as being produced by making a horizontal cylinder around the earth. The term transverse arises from the fact that the axis of the cylinder is perpendicular or transverse to the axis of rotation of the earth. In the UTM coordinate system, the earth is divided into 60 zones, each 6º of longitude in width, and the UTM projection is applied to each zone using its centerline as the principle meridian. The cylinder can be made so that it touches the earth's surface along one of the center-lines of one zone so that no point in a given zone is more than 3º from the location where earth distance is truly preserved. UTM projection coordinates are the standard for urban airshed modeling of atmospheric photochemistry. The required projection for this project was a UTM zone 15 projection, which has its central meridian 93 degrees West of the Prime Meridian.

Table 2.2: North Central Texas Council of Governments Land Use Categories

Anderson Level Three Classification		Anderson Level Four Classification	
111	Single Family	1111	Rural Density
		1112	Moderate Density
		1113	High Density Urban
112	Multi-Family		
113	Mobile Home Parks		
114	Group Quarters		
121	Office		
122	Retail	1221	Regional Mall (Area over 500,000 ft^2)
		1222	Major Shopping Center (Area between 500,000 and 50,000 ft^2)
		1223	General Shopping (Area less than 50,000 ft^2)
123	Institutional	1231	Churches
		1232	Education
		1233	Military Bases
124	Hotel/Motel		
131	Industrial		
141	Trans./Communication		
142	Roadways		
143	Utilities	1431	Water Facilities
		1432	Wastewater Facilities
		1433	Power Easements
144	Airports		
171	Parks and Recreation	1711	Golf Courses
		1712	Cemeteries
		1713	Public Parks
		1714	Amusement Parks
172	Landfill		
173	Construction		
181	Flood Control		
300	Vacant		
500	Water		

Texas Parks and Wildlife data on vegetation cover

The Texas Parks and Wildlife Department vegetation data were used to assign land cover over most of the rural sections of the study domain. Although the data have a lower spatial resolution than some of the USGS data, ground surveys described later in this report indicated that the land cover assignments were more accurate in the Texas Parks and Wildlife Department data.

The data set was completed by the Texas Parks and Wildlife Department in 1984. It is a compilation of ground-truthed satellite imagery, preexisting vegetation maps, and land resource, or land use, units. For the eastern two-thirds of the state, a computer technique was used to categorize the vegetation with ground-truth and Landsat data. The earth-satellite flights used for the data took place from 1975 to 1981. In the western portion of the state, the vegetation overstory was not as significant, and the classification of the vegetation was based on previously delineated land resource units from a survey conducted by the Bureau of Economic Geology (BEG) and the University of Texas (TP&W website, http://www.tpwd.state.tx.us/admin/veg/intro.html, 1996). The land resource units on the western side of the state were overlaid onto the Landsat vegetation data. Land classification boundaries were then erased when adjacent areas showed the same vegetation signatures. (Texas Parks and Wildlife, 1996) The land cover categories for this database are listed in Table 2.3.

Table 2.3: Texas Parks and Wildlife Department Vegetation Classifications

Land Use Code	**Description**	**Percent of Total Domain**
200	Crops	22.6
401	Ashe Juniper Parks/ Woods	1.2
402	Bluestem Grassland	7.4
403	Cottonwood - Hackberry – Saltcedar Brush/Woods	0.2
406	Elm – Hackberry	3.1
408	Live Oak - Ashe Juniper Parks	1.4
409	Live Oak - Ashe Juniper Woods	0.3
411	Live Oak-Mesquite - Ashe Juniper Parks	2.6
413	Mesquite-Lotebush Brush	3.2
414	Mesquite Brush	0.2
415	Oak - Mesquite - Juniper Parks/Woods	8.1
416	Other	7.8
418	Pine - Hardwood Forest (Loblolly Pine - Sweetgum) or Pine – Hardwood Forest (Shortleaf Pine - Post Oak - Southern Red Oak)	2.4
419	Post Oak Parks/Woods	2.4
420	Post Oak Woods, Forest and Grassland	18.3
421	Post Oak Woods/Forest	7.3
422	Silver - Bluestem - Texas Wintergrass Grassland	5.3
423	Urban	2.3
424	Willow Oak - Water Oak – Blackgum Forest	0.2
425	Water Oak - Elm - Hackberry Forest	1.4
500	Water	2.4

The land use classifications in this Texas Parks and Wildlife Department database distinguish between woods, parks, forests and grasslands. Operational definitions for these terms are provided in the TNRIS web site and are summarized in Table 2.4.

Table 2.4: Definitions of Forests, Grasslands, Parks and Woods Used in the Texas Parks and Wildlife Department Database.

Vegetation Types	**Definition as Given by the Texas Parks and Wildlife Dept.**
Grassland	Herbs (grasses, forbs, and grasslike plants) dominant; woody vegetation lacking or nearly so (generally less than 10% or less woody canopy coverage
Shrub	Individual woody plants generally less than 9 feet tall scattered throughout arid or semi-arid regions (less than 30% woody canopy coverage)
Brush	Woody plants mostly less than 9 feet tall dominant and growing as closely spaced individuals, clusters, or closed canopied stands (greater than 10% canopy cover)
Parks	Woody plants mostly equal to or greater than 9 feet tall generally dominant and growing as clusters, or as scattered individuals within continuous grass or forbs (11-70% woody canopy cover overall)
Woods	Woody plants mostly 9 to 30 feet tall with closed crowns or nearly so (71%-100% canopy cover); midstory usually lacking
Forest	Deciduous or evergreen trees dominant; mostly greater than 30 feet tall with closed crowns or nearly so (71-100% canopy cover); midstory generally apparent except in managed monoculture

The USDA-National Agricultural Statistics Service (NASS)

The Texas Parks and Wildlife mapping provides only one land cover category for cropland. In order to refine the leaf biomass density assignments for cropland, data on crop harvests in 1995 were obtained from the USDA-National Agricultural Statistics Service. The information at the USDA-NASS included acres planted, harvested, yield per acre, production, and commodity cropping practice for counties all over the United States. There were seven types of crops that were grown in the counties of the study domain according to this database. The data for these crops (corn, wheat, sunflowers, soybeans, sorghum, oats, and cotton) enabled a more detailed characterization of the emissions of the domain.

2.4 PROCESSED DATABASES NOT USED

Although data were obtained from Landsat, the USGS and other sources, these data were not used in the development of leaf biomass density estimates in this study. The rationale for not using these data is documented below.

The Landsat Thematic Mapper data were not used for the development of the leaf biomass assignments. The imagery is expensive, and it was decided that there was not enough added benefit to the database to offset the time and money it would take to develop the data into leaf biomass estimates. Furthermore, the image that was developed did not give enough information on the difference between the vegetation species of the area. The image could only roughly distinguish between herbaceous and arboreal vegetation. This information alone was not enough to assign accurate biogenic emission factors. With additional images taken at later dates, the species of vegetation might be better determined by their foliar cycles. Also ground truthing of the data might provide sufficient information about the vegetation types.

The USGS LULC data were not used because the LULC assignments did not provide enough detail on species distributions in the study domain. This problem is clear from the USGS LULC categories (Table 2.5). Forests, for example, are only categorized into deciduous, evergreen and mixed categories. This level of detail does not allow trees that release large quantities of hydrocarbons (e.g. oaks) to be separated from deciduous trees that have lower emission rates. The USGS LULC data also lack specificity. As shown in Figure 2.3, large sections of the study domain have land cover classifications of “crop and pasture land” and “herbaceous rangeland." These are not particularly useful classifications for estimating leaf biomass densities.

Table 2.5: USGS Land Use/ Land Cover Categories

Code	Classification	Percent of Total Area
11	Residential	2.28
12	Commercial and Services	0.58
13	Industrial	0.10
14	Transportation, Communication, Utilities	0.46
15	Industrial and Commercial Complexes	0.07
16	Mixed Urban or Built-up Land	0.05
17	Other Urban or Built-up Land	0.27
21	Cropland and Pasture	58.88
22	Orchards, Groves, Vineyards, Nurseries	0.06
23	Confined Feeding Operations	0.02
24	Other Agricultural Land	0.04
31	Herbaceous Rangeland	3.14
32	Shrub and Brush Rangeland	4.99
33	Mixed Rangeland	10.31
41	Deciduous Forest Land	11.28
42	Evergreen forest land	0.98
43	Mixed Forest Land	3.94
51	Streams and Canals	0.04
52	Lakes	0.01
53	Reservoirs	1.61
61	Forested Wet Land	0.18
62	Nonforested Wet Land	0.14
72	Beaches	0.00
73	Sandy Areas not Beaches	0.01
75	Strip Mines, Quarries, Gravel Pits	0.12
76	Transitional Areas	0.46

The USGS Land Cover Characteristics database was not used in the final vegetation database constructed for the database. Although the USGS LCC data were at a higher resolution than that of the Texas Parks and Wildlife Department vegetation coverage, the classifications used in the USGS LCC were not as valuable for biogenic hydrocarbon emissions estimation. These classifications are shown in Table 2.6. More discussion on this topic is presented in the results chapter of this document.

Table 2.6: U.S. Geological Survey Land Cover Characteristics

USGS Land Cover Code	Description	Percent of Total Area
2	Cropland (Small Grains) with Grasslands	0.001
12	Cropland (Corn and Soybeans)	0.001
13	Cropland (Small Grains, Hay, Pasture) with Wetlands	1.179
14	Cropland (Mixed Row Crops) with Woodland	0.040
17	Cropland (Winter Wheat)	0.378
19	Cropland (Cotton, Soybeans, Rice)	0.014
21	Cropland (Cultivated Grasses) with Woodlands	0.020
23	Cropland (Corn, Soybeans, Cotton, Rice) with Woodlands	22.867
32	Irrigated Agriculture	< 0.001
43	Cropland (Corn, Small Grains)/ Deciduous Forest (Oak, Hickory)	0.041
44	Cropland (Corn, Soybeans, Pasture)/ Woodland (Oak, Hickory)	0.008
47	Cropland (Corn, Cotton, Soybeans)/ Evergreen Needleleaf Forest	0.001
50	Grassland/ Cropland (Wheat, Corn) Mosaic	0.031
52	Cropland (Small Grains, Row Crops)/ Grassland	0.074
53	Cropland (Corn, Sorghum, Small Grains) Grassland Mosaic	0.103
54	Cropland/ Grassland	3.603
55	Cropland (Corn, Cotton, Sorghum, Pasture)/ Grassland Mosaic	9.816
56	Cropland (Pasture)/ Grassland Mosaic	0.006
58	Grassland (Short-Midgrass Prairie)	0.018
59	Grassland (Short Grass Prairie)	0.006
62	Grassland	0.011
63	Grassland (Warm Season Grasses)	0.023
64	Grassland with Cropland (Small Grains, Pasture)	2.157
65	Mixed Rangeland (Grassland and Shrubland)	0.182
67	Grassland with Cropland	34.026

68	Grassland (Tall Grass Prairie)	0.014
70	Savanna	0.003
72	Mixed Rangeland (Needlegrass, Big Sage, Rabbitbrush)	0.001
74	Mixed Rangeland (Shrubs and Grasses)	0.250
75	Mixed Rangeland (Grassland, Shrubland) with Crops, Fallow)	0.004
83	Tall/ Low Shrubs, Tundra, Spruce	0.014
94	Grassland/ Forest	16.514
98	Open Mixed Forest (Aspen, Birch, White Spruce, Black Spruce)	0.003
104	Deciduous Woodlands (Aspen)/ Shrublands (Mountain Mahogany)	0.055
105	Mixed Forest (Aspen, Birch, Spruce, Balsam Fir)	3.306
112	Semi-Deciduous Tropical Forest	0.003
131	Needleleaf Boreal Forest (Black and White Spruce, Tamarack),	0.023
135	Needleleaf Forest (Douglas Fir, Spruce, Western Red Cedar)	0.005
138	Needleleaf Forest (Hemlock, Spruce, Douglas Fir)	0.001
140	Evergreen Needleleaf Forest (Douglas Fir, Lodgepole Pine, Larch)	0.060
147	Needleleaf Forest (Ponderosa, Lodgepole and White Pine, Douglas Fir)	0.873
153	Evergreen Needleleaf Forest (Spruce, Balsam Fir, Eastern White Pine, Eastern Hemlock)	0.001
154	Needleleaf Forest (Douglas Fir) with Mixed Hardwoods	1.167
157	Needleleaf Forest (Douglas Fir)	0.022
164	Open Mixed Forest (Spruce, Aspen)	0.004
170	Mixed Forest (Pine, Oak)	0.093
171	Mixed Forest (Oak, Pine Species)	0.016
183	Ponderosa/Lodgepole Pine Woodland	0.032
186	Northern Mixed Forest (Maple, Beech, Birch, Pine)	0.843
205	Water	2.086

2.5 Obtaining and Processing the data in a GIS

United States Geological Survey LULC data

The USGS LULC data are available from a few locations. The data for the domain of interest was obtained from the USGS ftp site located at edcftp.cr.usgs.gov. The name of the 1° latitude by 2° longitude files for the domain can be found on the USGS 1 to 250,000 scale map index webpage at http://edcwww.cr.usgs.gov/Webglis/glisbin/searchchoiceftp.pl?dataset_name=1_250_LULC. A user's guide for the information is on the webpage found at http://edcwww.cr.usgs.gov/glis/hyper/guide/1_250_lulc. The desired USGS LULC map sections for the North-Central Texas domain are shown in Figure 2.1. The 1° latitude by 2° longitudinal map sheets needed to describe the entire domain were those for Austin, Abilene, Beaumont, Lawton, Wichita Falls, Sherman, Texarkana, Dallas, Tyler, Palestine, Waco, and Brownwood.

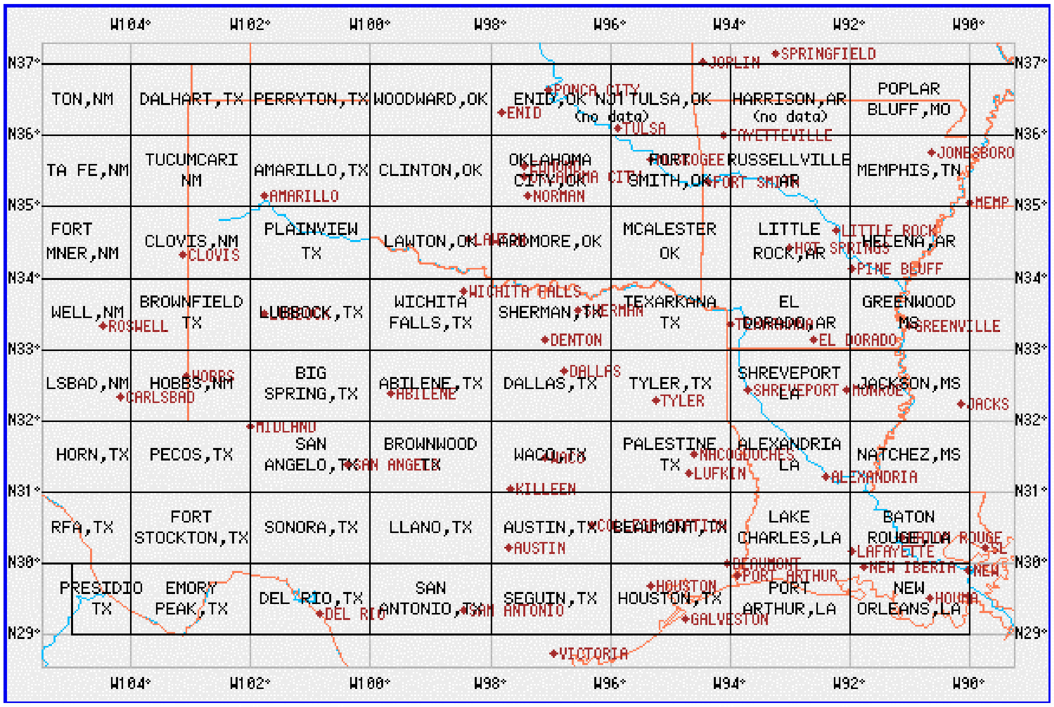


Figure 2.1: USGS LULC index map showing the map sections needed for the North-Central Texas study domain.

The first step was to ftp the needed files from their source to the working directory. The ftp site used was at edcftp.cr.usgs.gov. The "anonymous" ftp method was used by typing anonymous at the Name prompt. A complete e-mail address was then entered at the Password prompt. The directory was changed (**cd**) to the "pub/data/LULC/250K" subdirectory to access 1:250,000-scale LULC data. The file transfer mode was set to binary by typing the word **binary**. The **get** and **mget** commands were then used to download the 00README or data files. The 00README file located under "/pub/data/LULC/250K" contained an explanation of the directory structure. The map sections were then downloaded in Arc/Info export format (files with the suffix .e00).

The Palestine coverage was not available from the USGS site. Administrators mention the problem is processing related. The file for this map sheet was obtained from the Texas Natural Resource Information System (TNRIS) site at http://www.tnris.state.tx.us/gispage.html. All of the USGS LULC files for the state of Texas are located in this directory in Arc/Info export format. In both cases the files needed to be unzipped before they could be used. This was done on the unix machines using the **gunzip** command:

& **gunzip dallas.e00.gz**

Once the files were downloaded and unzipped, Arc was used to import the files into usable coverages using the following command:

Arc: **import cover <dallas.e00> <dallas>**

The images then needed to be cleaned and the topology needed to be built. By cleaning the coverage, the program was checking for any arc intersections that did not have nodes or unconnected (dangling) arcs. The clean and build commands were used for this.

Arc: **clean <dallas>**

Arc: **build <dallas>**

These three commands were then repeated for each file that was downloaded and unzipped. Upon looking at each coverage, lines separating like land use areas can be seen which were used in creating the coverages. These lines were eliminated by using the dissolve command. Dissolve removed any lines separating polygons that had the same attribute specification. Lu_code was the attribute specified in the dissolve command.

Arc: dissolve dallas ddallas lu_code poly

(Where the extra "d" in ddallas was the nomenclature used to represent the dissolved coverage.)

The individual USGS LULC files could then used in Arcview. Figure 2.2 shows an example of one of the individual coverages. The legend for this USGS LULC data is shown in Table 2.7 on page 40.

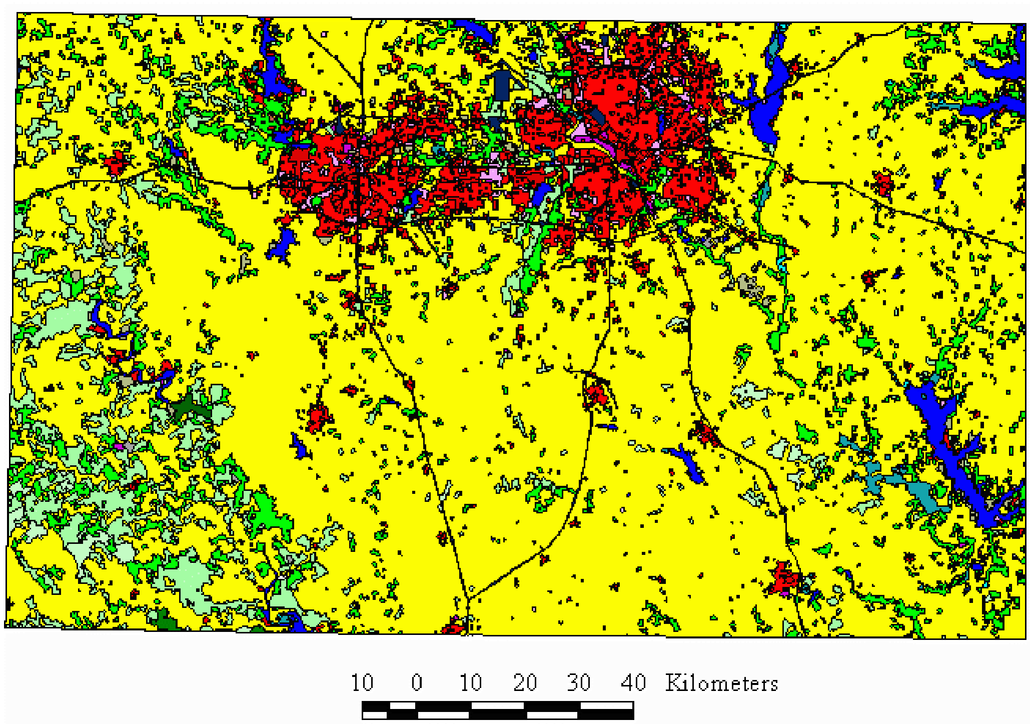


Figure 2.2: A one by two degree USGS Land Use/ Land Cover file (Dallas area)

After the USGS LULC files for the study domain were usable, they needed to be combined into a single coverage. This was done with the use of Arc/Info. In Arc, the **mapjoin** command was used to combine all the coverages. Unfortunately, this command is not always successful. Adjacent coverages can sometimes have inconsistencies in the lines and nodes between them, and they could need to be edited. The required editing could be done before or after the pieces are joined so long as there are no major discrepancies that do not allow for joining. For this database the editing was done after the coverages were joined. They were joined in Arc using the following command:

Arc: **mapjoin <combined>**

(where combined is the name of the new combined coverage)

Enter the 1^{st} coverage: **abilene**

Enter the 2^{nd} coverage: **dallas**

Enter the 3^{rd} coverage: **wichitafalls**

Enter the 4^{th} coverage: **waco**

Enter the 5^{th} coverage: **lawton**

Enter the 6^{th} coverage: **texarkana**

Enter the 7^{th} coverage: **brownwood**

Enter the 8^{th} coverage: **palestine**

Enter the 9^{th} coverage: **tyler**

Enter the 10^{th} coverage: **sherman**

Enter the 11^{th} coverage: **austin**

Enter the 12^{th} coverage: **beaumont**

Done entering coverage names?(y/n) : **y**

Do you wish to use the above coverages?(y/n) : **y**

If the composite coverage had not worked, then some editing would have been done before the smaller coverages were joined. For this project the mapjoin was successful. However, the map sheets did not align at the edges. This resulted in gaps between map sheets, and these gaps resulted in area polygons with no attributes. The gaps were often too large for the **dissolve** command to eliminate, and they were removed manually using ArcEdit. For this work the image was shown in ArcEdit using the **gridpaint** command. The edit work was done manually in ArcEdit.

To edit a coverage in ArcEdit, the coverage and the feature to be edited (node, arc, or polygon) were specified.

Arc: **ae** (started ArcEdit)

Edit: **display 9999** (made a display window active on the unix machine)

Edit: **ec** <**combined**> (the edit coverage was set to combined)

Edit: **de** <**combined**> **arc** (specified that arcs were to be displayed)

Edit: **de** <**combined> node** (specified that nodes were to be displayed)

Edit: **draw** (the nodes and arcs of the coverage were drawn)

Edit: **edit feature nodes** or **ef arc** (determined which features were to be edited)

The details of the coverage's nodes and arcs were then magnified or reduced by using zoom commands. While the image was active, the 'zoom in' function could be done with **Ctrl e**, the 'zoom out' was **Ctrl x**, and the full extent could be seen with **Ctrl f**. Features could be chosen with the **sel many** (select many) command. Features could then be deleted with the **delete** command or moved with the **move** command. If features needed to be added, they were specified with the **ef** (edit feature) command and then added with the **add** or **add many** commands. Any mistakes were corrected with the **oops** command which cancelled the last operation. The following simple guidelines were fundamental; to have an arc, there needed to be a node at each end; to have a polygon, there needed to be a closed set of arcs. Once the editing work was satisfactory the work was saved with the **save** command.

Finally, the **dissolve** command was used in Arc to ensure that no lines separated areas with the same land use code.

Arc: **dissolve combined dcombined lu_code poly**

After the coverage was sufficiently edited, it was projected into the proper coordinate system. As stated before, the database was to be in UTM zone 15 coordinates. Although the domain is in zone 14, zone 15 coordinates ensured that the database was compatible with the earlier Gulf-Coast biogenics study done by the TNRCC for the Houston area.

The USGS LULC data was originally in a UTM projection, but the data were not in UTM zone 15 coordinates, so the files were projected into the UTM zone 15 coordinate system. There are two ways to project a coverage into another coordinate system in Arc/Info. The projection can either done manually or with the use of a projection file. The manual projection is done with a computer prompting; the projection file method is a one-line command where the projection parameters are pre-written in a text file. The projection file approach is recommended because this allows the operator to go back and review the file if the results are not satisfactory. A manual series of commands are more difficult to review if the projection is not successful.

While in Arc, the **describe** command is used to see what projection the coverage is in.

Arc: **describe <file>**

This command displays the information known about the coverage, including the current coordinate projection. When the program gives a message stating that the projection has not been defined, the projection of the coverage needs to be defined using the **projectdefine** command. It is a challenging task to determine in what coordinate system a coverage is located without any metadata information about the coverage. If no metadata are available, trial and error is used to try and fit the coverage into an area with a known coordinate system and known relation to the original coverage.

When the **describe** command successfully defines the projection of the coverage, the projection can be completed manually or with a projection file with the following information:

-- output

-- projection utm

-- units ____ (either **meters** or **0.001** for kilometers)

-- zone ____ (UTM zone **15** desired)

-- parameters

-- end

A spheroid or datum command is also required if the input data of the coverage includes spheroid or datum information.

The projection command was as follows.

Arc: **project cover <input name> <output name>** {**projection file**} (if used. Some are given in appendix B)

The **describe** command was used with a manual input for most of this work.

Arc: **describe <input file>**

Arc: **project cover <input name> <output name>**

****

*	The INPUT projection has been defined.	*

****

Use OUTPUT to define the output projection and END

to finish.

Project: **output**

Project: **projection utm**

Project: **units ____**

Project: **zone ____**

Project: **parameters**

Project: **end**

Once the coverage was edited in its combined state and projected into the UTM zone 15 coordinates, it was clipped so that only the land use for the 37 county region was given. This was done using the **clip** command in Arc. The **clip** command was analogous to a cookie cutter in that another coverage was used to define the desired area to be kept, and the rest of the area was cut away. For this coverage the **clip** command was used with the county file being the desired area.

Arc: **clip <dcombined> <county> <finalcombined>**

Once again the **clean** and **build** commands were used.

Arc: **clean <finalcombined>**

Arc: **build <finalcombined> poly**

Finally the **dissolve** command was used to eliminate any arcs (lines) that were separating adjacent areas with the same land use code.

Arc: **dissolve cover <finalcombined> lucode**

(In the above command, lucode is telling the program what attribute or field in the attribute table, of the coverage to use in the dissolving process)

The final clipped USGS LULC coverage is shown in Figure 2.3. Table 2.7 shows the legend for the USGS LULC data.

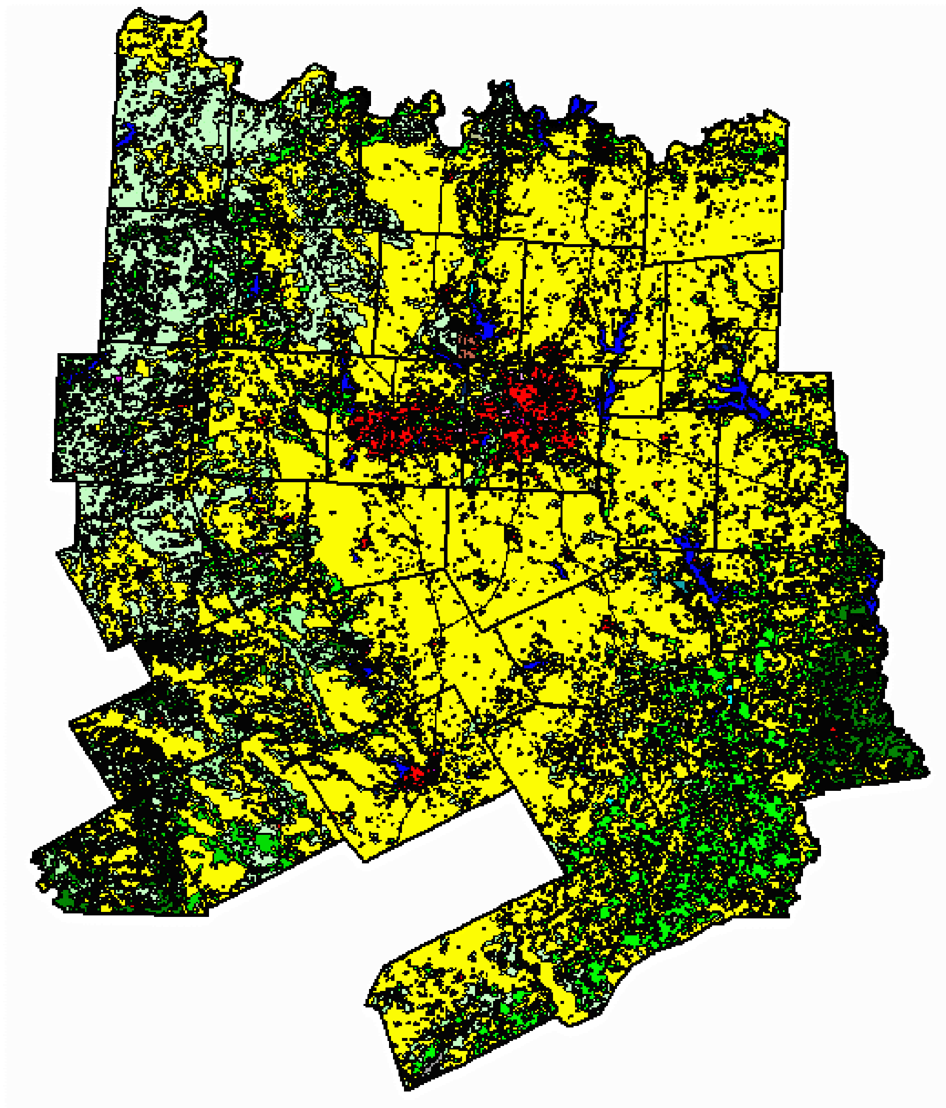


Figure 2.3: Final joined, dissolved, and edited USGS LULC coverage for the 37 county study domain.

Table 2.7: USGS Land Use Land Cover Legend

Residential
Commercial and Services
Industrial
Transportation, Communication, and Utilities
Industrial and Commercial Complexes
Mixed Urban
Other Urban
Crop and Pasture Land
Orchards, Vineyards, and Nurseries
Confined Feeding Operations
Other Agricultural Land
Herbaceous Rangeland
Shrub and Brush Rangeland
Mixed Rangeland
Diciduous Forest
Evergreen Forest
Mixed Forest
Lakes
Reservoirs
Forested Wetland
Nonforested Wetland
Non-beach Sandy Areas
Strip Mines and Quarries
Transitional Areas

North-Central Texas Council of Governments LULC

The North-Central Texas Council of Government's Land Use Land Cover database was obtained from the NCTCOG through the TNRCC. For this project the Anderson level three classifications were used. The attribute table for these classifications is shown in Table 2.7.

Table 2.8: North-Central Texas Council of Governments classifications

Land Use Code	Description	Percent of Domain Area
111	Single Family	13.95
112	Multi-Family	1.02
113	Mobile Home Parks	1.38
114	Group Quarters	0.08
121	Office	0.39
122	Retail	1.53
123	Institutional	1.37
124	Hotel/Motel	0.04
131	Industrial	2.60
141	Trans./Communication	0.18
142	Roadways	1.33
143	Utilities	0.74
144	Airports	0.73
171	Parks and Recreation	2.69
172	Landfill	0.10
173	Construction	0.97
181	Flood Control	0.20
300	Vacant	66.47
500	Water	4.22

This coverage was originally in the UTM zone 14 coordinate system. It was projected into UTM zone 15 coordinates using the previously described manual projecting procedures.

Arc: **describe <nctcoglu>**

Arc: **project cover <nctcoglu> <nctcog15>**

****

*   The INPUT projection has been defined.   *

****

Use OUTPUT to define the output projection and END

to finish.

Project: **output**

Project: **projection utm**

Project: **zone 15**

Project: **units 0.001**

Project: **parameters**

Project: **end**

The projected coverage was then cleaned and the topology was built.

Arc: **clean <nctcog15>**

Arc: **build <nctcog15> poly**

The final NCTCOG land use database projected in UTM zone 15 coordinates is shown in Figure 2.4. The legend for the North Central texas Council of Governments land use data is shown in Table 2.9.

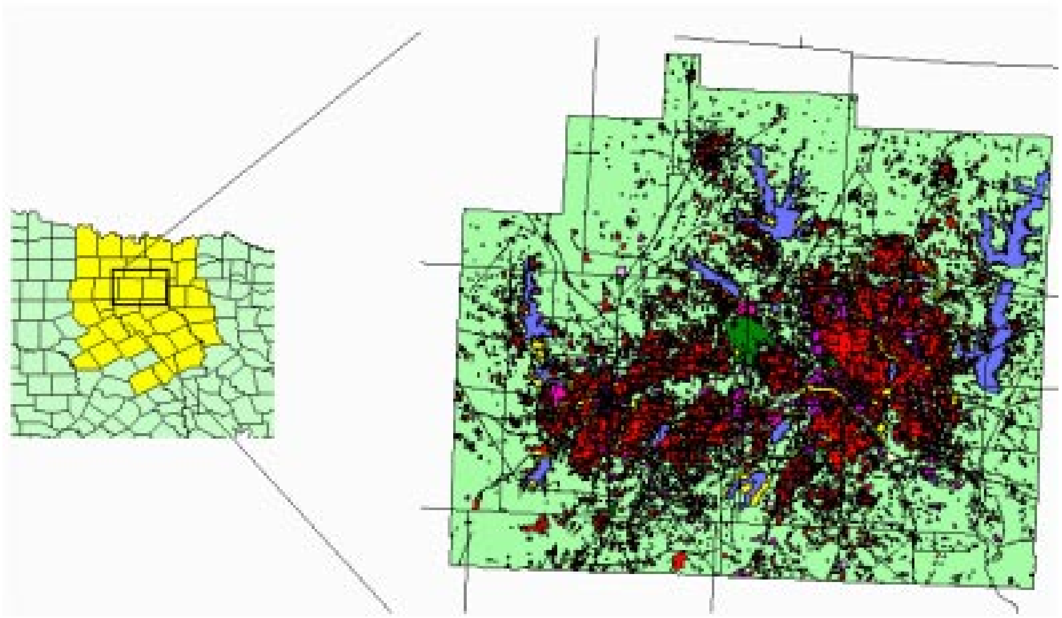


Figure 2.4: NCTCOG projected into UTM 15 coordinates

Table 2.9: Legend for North Central Texas Council of Governments Data

Residential
Office
Retail
Institutional
Hotel-Motel
Industrial
Transportation and Communication
Roadway
Utilities
Airport
Parks and Recreation
Landfill
Construction
Flood Control
Vacant
Water

USGS Land Cover Characteristics Data Base

The USGS LCC database is one of the most comprehensive sources of vegetation data that can be utilized in a GIS. These data are part of a project to develop one kilometer vegetation data for the entire world. The data can be found at the web site http://edcwww.cr.usgs.gov/landdaac/glcc/glcc.html. Although this data base is public domain, the conversion of the data to an Arc/Info usable format is not straight-forward. The following process was used to make the vegetation data useable in the Arc/Info and Arcview geographic information systems.

Arc/Info directly supports the Lambert Azimuthal Equal Area projection so it was more efficient to obtain and project the data in this projection. The file was downloaded off the web directly from the North American Land Cover Lambert Azimuthal Equal Area Projection File Listing found at http://edcwww.cr.usgs.gov/landdaac/glcc/tablambert_na.html. The file was also downloadable via ftp from ftp://edcftp.cr.usgs.gov in the directory /pub/data/glcc/na. The files were on the order of 83 megabytes so the available gzipped files were used.

The files were in .img format which could be viewed with some types of cartographic software. To use the file in Arc/Info, the data were converted from the .img format to a grid. First the file was unzipped. The unzipping command at the UNIX environment was as follows.

& **gunzip <lcc.img>**

Next the file was given a "bil" extension.

mv <lcc.img> <lcc.bil>

A header file was then made for the .bil file. This header file contained information on the size and location of the data set. The file was made in a text editor, and it needed to be saved in the same directory as the 'bil' file and needed the same name as the bil file.

lcc.hdr

The metadata needed for the header files was found at http://edcwww.cr.usgs.gov/landdaac/glcc/nadoc.html under geometric characteristics. The header file created for the Seasonal Land Cover data set was as follows. The words in parenthesis are only for explanation.

Header Line	**Explanation**
nrows 8996	(Number of rows in the data set)
ncols 9223	(Number of columns in the data set)
nbits 8	(The bit size of the image)
ulxmap 4486550	(The x-coordinate of the upper left corner of the image)
ulymap 4479869	(The y-coordinate of the upper left corner of the image)
xdim 1000	(Number of meters in x direction for each pixel)
ydim 1000	(Number of meters in y direction for each pixel)

Two files were now in the working directory; <lcc.img> and <lcc.hdr>. The file was then converted into a grid file. The .bil file was converted to a grid file using the **imagegrid** command in Arc.

Arc: **imagegrid <lcc.bil> <lcc>**

Arc was then used to describe the current projection's coordinate system using the **projectdefine** command.

Arc: **projectdefine <lcc>**

Project: **projection lambert_azimuthal**

Project: **parameters**

radius of sphere of reference [0.00000]: **6370997**

longitude of center of projection [0 0 0.000]: **-100 0 0.000**

latitude of center of projection [0 0 0.000]: **50 0 0.000**

false easting (meters) [0.00000]: **0.00000**

false northing [0.00000]: **0.00000**

All of the above projection parameters for this file were taken from the Lambert Azimuthal Equal Area Projection Parameters found at http://edcwww.cr.usgs.gov/landdaac/glcc/nadoc.html#lamb.

Up to this point the data in the working directory was for the entire North American continent. Only a small section of this file was needed, and clipping away the unwanted portions of the coverage would save processing time. The first step was to identify the area that was to be isolated on the grid of North America. This was done in the **grid** subset of commands in Arc.

Arc: **grid**

Grid: **display 9999** (opened a viewing window)

Grid: **mapextent lcc** (set the range of the window to that of the file)

Grid: **gridpaint <lcc>** (painted the file in color)

Gridpaint lcc value linear nowrap gray was used for a gray image, and **clear** was used to erase any image. Next, the area of interest was selected from the file. Water landmarks were used for orientation of the file. The AVHRR value read for all water was the same and is shown in black with this display. The lakes and the shoreline of the continent were used to locate the desired domain. Any extra area that could be removed later was left when uncertainties existed. The user interface on the view window allowed for zooming in and out when more detail was desired. The region of interest was then separated.

Grid: **setwindow *** (This command allowed for a drawing of a rectangle which was the area of the new file. This was done using the cursor in the image window and clicking the mouse button to make two opposite corners of the selection rectangle.)

Grid: **lcc2=lcc** (Created and named the smaller grid)

Grid: **clear**

Grid: **setwindow lcc2**

Grid: **gridpaint lcc2**

A much smaller and more convenient file was now shown. The process could be repeated as needed (The **clip** command in Arc was also used later).

The grid file that was in the working directory at this point did not have an attribute table. The only data that this file contained were the numerical values that represented the land cover type for each separate square kilometer. There were two options for completing in the attribute table for the file. One option was to use the Info subprogram of ArcInfo. The alternative used for this project was to use Arcview methodologies. For the Arcview method, the grid was converted to a coverage (using Arc), and the additional information was added into the coverage's attribute table in Arcview.

Arc: **gridpoly lcc2 lcccoverage**

Arc: **clean lcccoverage**

Lcccoverage was then brought into Arcview, and the attribute table was modified. The only attribute of the coverage at this point was the gridcell value. Another table was created and given the gridcell values in the first column and the corresponding classification descriptions (shown in Table 2.6) in the second. This was done using the following method. First, **start editing** under the table menu was selected. **Add field** under the edit menu was used to add a new column to this second table. At the prompt, the name of the field, type of attribute, and size of the column were assigned. The descriptions were determined by looking at the legend on the website for the file being used. **Add rows** was selected for data entry into the table. The " I " icon was chosen over the table, and then the places to enter data were selected. Once the table was completed **stop editing** under the table menu was selected. After the second table was completed with a field for the cell values and a field for the descriptions, it was joined to the attribute table using the **link** command under the table menu. This process involved selecting the value field of the second table and then selecting the same field on the attribute table. Next, the **link** command was selected, and a new column was created in the attribute table with the corresponding classification for every value. Another option using Arcview could have been just to use the attribute table with only the value field in it and then design and store a legend with the corresponding descriptions in it.

The coverage was then projected into the required UTM 15 coordinate system. A projection file was used for this projection with the original coordinate system being the previously assigned lambert azimuthal projection. The Lambert to UTM projection file lamutm15.prj is given in Appendix A.

Arc: **project cover lcccoverage lccveg lamutm15.prj**

The coverage was then cleaned and the topology was built again.

Arc: **clean lccveg**

Arc: **build lccveg poly**

Finally, the land cover characteristics coverage was clipped down to the 37 county study region. This was done the same way that the USGS LULC coverage was reduced. The **clip** command was executed using the county file to define the desired area.

Arc: **clip <lccveg> <county> <lccvegcl>**

Once again the clean and build commands were used

Arc: **clean lccvegcl**

Arc: **build lccvegcl poly**

The final clipped version of the USGS LCC data is shown in Figure 2.5. The legend for the Figure 2.5 is given in Table 2.10.

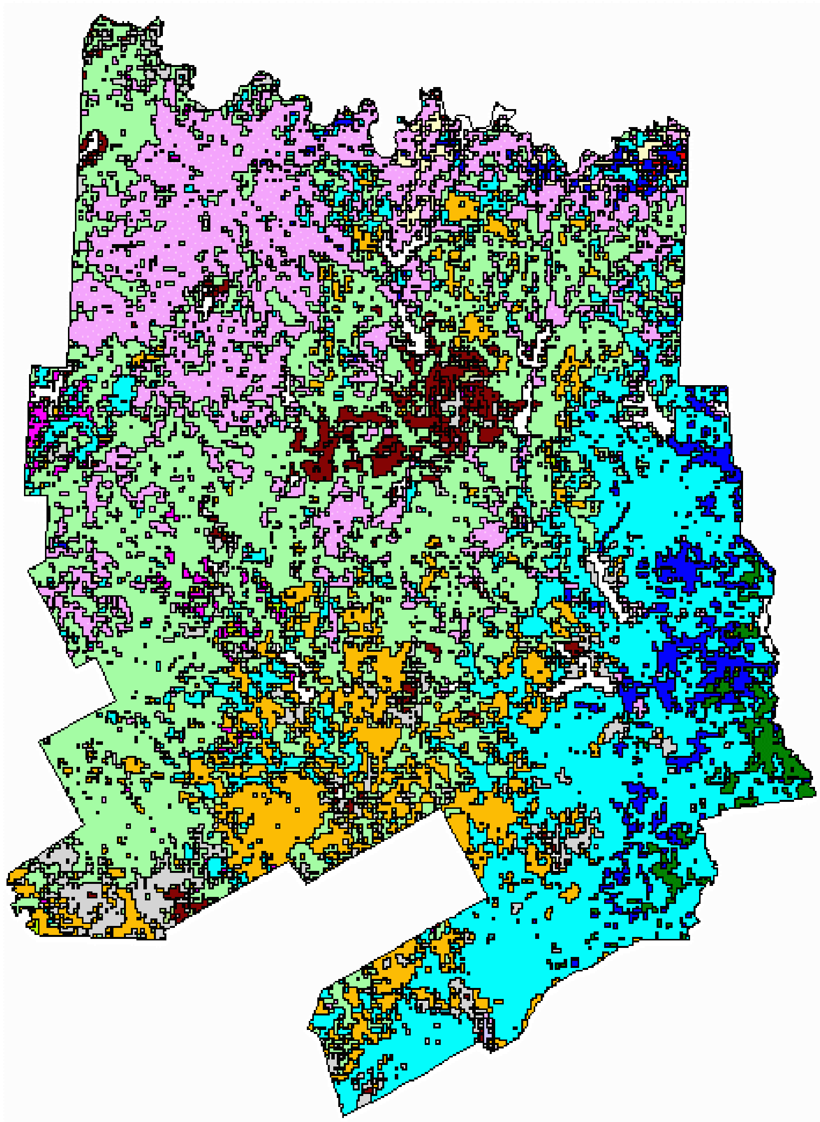
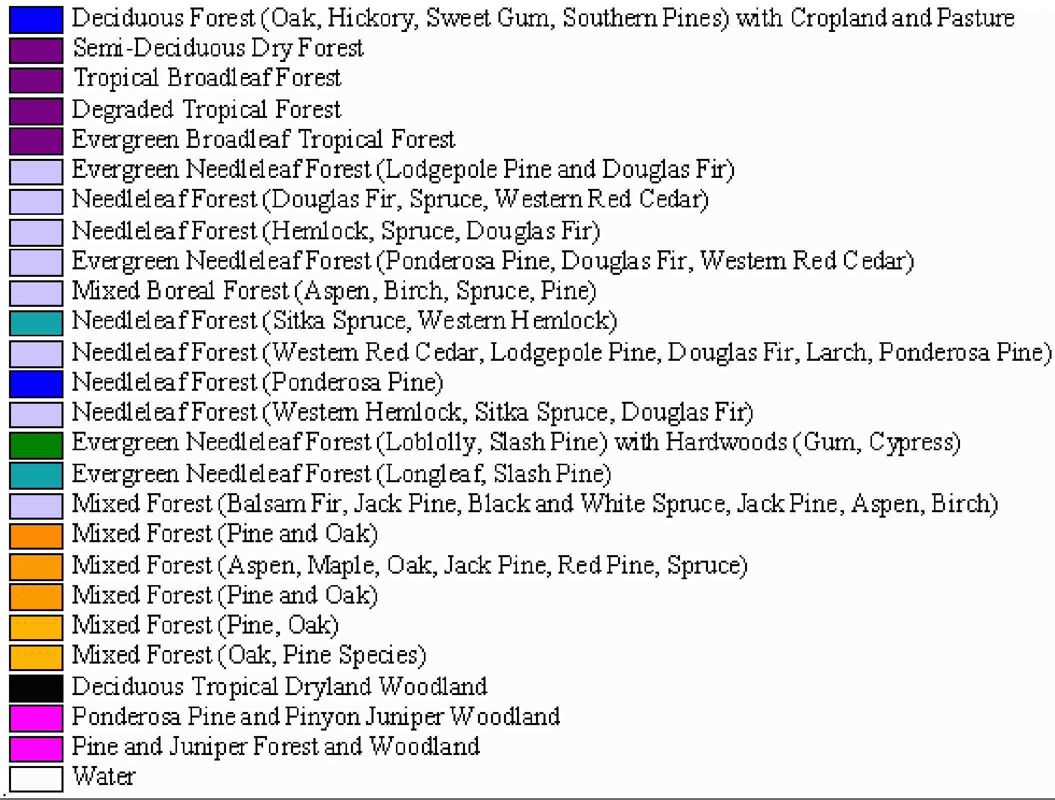


Figure 2.5: The USGS LCC data coverage for the 37 county study domain.

Table 2.10: USGS Land Cover Characteristics Legend

Cropland (Small Grains) with Grasslands
Cropland
Cropland (Corn and Soybeans)
Cropland (Corn and Soybeans)
Cropland (Small Grains, Hay, Pasture) with Wetlands
Cropland (Mixed Row Crops) with Woodland
Cropland (Winter Wheat)
Cropland (Cotton, Soybeans, Rice)
Cropland (Cultivated Grasses) with Woodland
Cropland (Corn, Soybeans, Cotton, Rice) with Woodlands
Irrigated Agriculture
Cropland (Corn, Small Grains)/Deciduous Forest (Oak, Hickory) Mosaic
Cropland (Corn, Soybeans, Pasture)/Woodland (Oak, Hickory) Mosaic
Cropland(corn, Cotton, Soybeans)/Evergreen Needleleaf Forest (Slash Pine) Mosaic
Cropland (Small Grains, Pasture)/Grassland Mosaic
Grassland/Cropland (Wheat, Corn) Mosaic
Cropland (Small Grains, Row Crops)/Grassland
Cropland (Corn, Sorghum, Small Grains)/Grassland Mosaic
Cropland/Grassland
Cropland (Corn, Cotton, Sorghum, Pasture)/Grassland Mosaic
Cropland (Pasture)/Grassland Mosaic
Grassland (Short mid Grass Prairie)
Grassland (Short Grass Prairie)
Grassland
Grassland (Warm Season Grasses)
Grassland with Cropland (Small Grains, Pasture)
Mixed Rangeland (Grassland and Shrubland)
Grassland with Cropland
Grassland (Tall Grass Prairie)
Savanna
Mixed Rangeland (Big Sage, Rabbitbrush, Needlegrass)
Mixed Rangeland (Saltbush, Sand Sage, Rabbitbrush)
Mixed Rangeland (Shrubs and Grasses)
Mixed Rangeland (Grassland, Shrubland) with Crops, Fallow
Desert Shrublands (Creosote, Saltbush, Sand Sage) Sonoran
Desert Shrubland/Grassland (Creosote, Saltbush, Mesquite, Sand Sage)
Desert Shrubland (Creosote, Saltbush, Mesquite, Cactus) with Grasses
Grassland/Woodland (Oak) Mosaic with Cropland
Open Needleleaf Forest (Ponderosa Pine and Lodgepole Pine)
Deciduous Forest (Maple, Beech, Birch, Oak, Hickory) with Pasture



Texas Parks and Wildlife Coverage

The Texas Parks and Wildlife Department vegetation coverage is available at the Texas Natural Resource Information Service URL http://www.tnris.state.tx.us/. These data were accessed by going to the GIS Data file and then to the V folder under A-Z. The data were filed under vegetation. The TP&WD data were available there in Arc/Info export format (.e00 suffix). The database described as vegcov.e00.gz was downloaded for this project. Once the file was downloaded, the file was gunzipped using the previously described method. The data was then imported into Arc/Info by using the import command in Arc.

& **gunzip vegcov.e00.gz**

Arc: **import cover vegcov.e00 vegcov**

Next, the coverage was projected into UTM 15 coordinates. The file was originally in a Lambert projection. Table 2.11 gives the coordinate description of the TP&WD data available at the TNRIS. These specific Lambert coordinates are called the Texas State Mapping System coordinates.

Table 2.11: Coordinate System Description of TP&WD data at TNRIS

Coordinate System Description	
Projection	LAMBERT
Units	METERS
Spheroid	GRS1980
Parameters:	
1st standard parallel	34 55 0.000
2nd standard parallel	27 25 0.000
Central meridian	-100 0 0.00
Latitude of projection's origin	31 10 0.000
False easting (meters)	1000000.00000
False northing (meters)	1000000.00000

To project into the Utm zone 15 coordinates, the manual projection method was used.

Arc: **Describe vegcov**

Arc: **project cover vegcov vegcovutm**

****

*   The INPUT projection has been defined.   *

****

Use OUTPUT to define the output projection and END

to finish.

Project: **output**

Project: **projection utm**

Project: **units 0.001**

Project: **zone 15**

Project: **parameters**

Project: **end**

This projected coverage was then cleaned and the topology was built.

Arc: **clean vegcovutm**

Arc: build vegcovutm

The TP&WD vegetation coverage then needed to be clipped to fit the domain of the study. The clipping was done as before using the 37 counties of the study as the clipping coverage. The clipping command used in Arc was as follows.

Arc: **clip cover vegcovutm counties vegcovcl**

In the above command the statement vegcovcl refers to the clipped vegetation coverage or the vegetation coverage for only the counties for the domain.

Again the coverage was cleaned and the topology was built.

Arc: **clean vegcovcl**

Arc: **build vegcovcl**

Figure 2.6 shows the TP&W coverage for the 37 counties in the study domain.

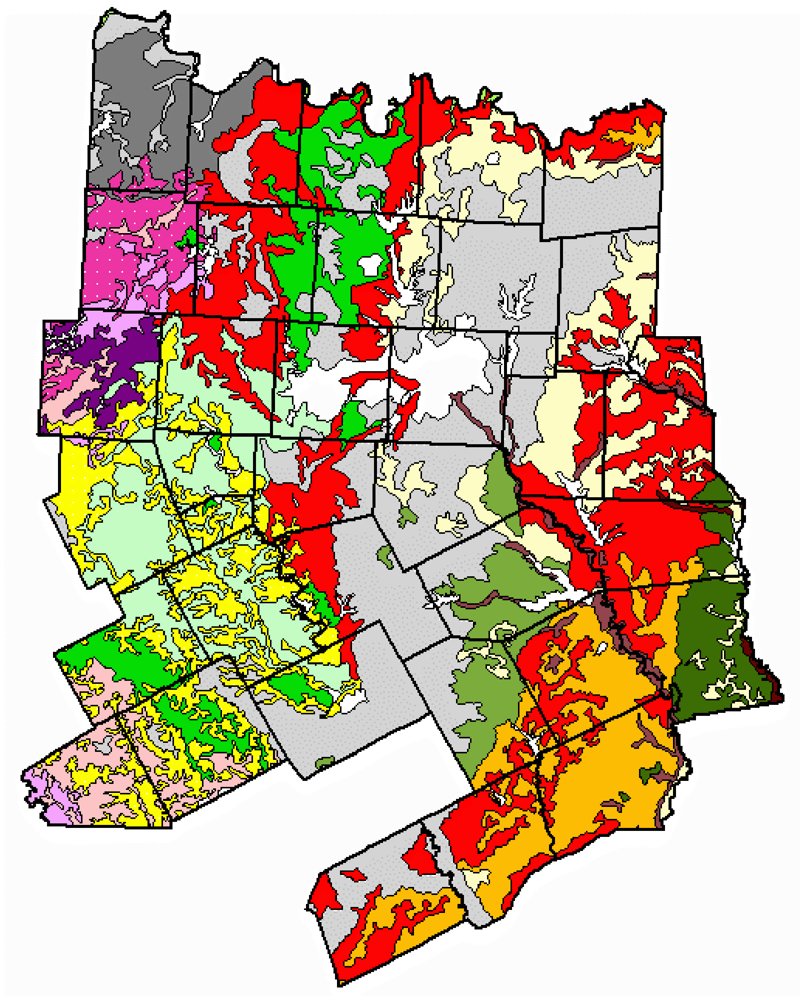


Figure 2.6: Texas Parks and Wildlife Department Vegetation Coverage for the 37 county domain

Table 2.12: Legend for Texas Parks and Wildlife Department vegetation data

USDA National Agricultural Statistics Service

The *USDA-National Agricultural Statistics Service (NASS)* provides data on crop harvests by county. The National Agricultural Statistics Service home page can be found at http://www.usda.gov/nass/nasshome.htm. Agricultural data are available for most crop producing counties in the United States.

The USDA crop statistics for 1995 and 1996 are found at the website at http://www.mannlib.cornell.edu/data-sets/crops/9X100/E/. This site has USDA-NASS Crops County Data files along with an information file called Readme. The data files are by crop type for the entire nation. The crop files having Texas data were barley, cotton, rice, sorghum, soybeans, sugarbeats, sugar cane, winter wheat, oats, corn, sunflowers, and total wheat. Within each crop file several statistics were given for each individual county. Each county was identified by its FIPS code (given in Table 1.2). The data given and format of the data are shown in Table 2.13.

Table 2.13: Data Format given for each county in USDA-NASS database

Column	Data
1-5	Survey Code
6-7	State FIPS Code
8-9	NASS Agricultural Statistics District (ASD)
10-12	FIPS Counry Code
13-14	Year Keyed
15-16	Crop Year
17-22	Commodity Code
23-26	Commodity Cropping Practice
27-34	Acres Planted All Purpose
35-42	Net Seeded Acres or Acres Planted for Harvest
43-50	Acres Harvested
51-54	Yield per Net Seeded or Yield per Planted Acre or Yield per Acre Planted for Grain
55-58	Yield Per Harvested Acre
59-68	Production
69-78	Sucrose Content for Sugarcane and Sugarbeets

For this study, each of the crop files in the database were downloaded and uncompressed. Each file was then examined to see if it contained records for Texas's state FIPS code. The ones that did have Texas data, FIPS code 48, were then analyzed for the specific counties of the domain. The amount of each crop planted in each county was calculated by using the data found under "net seeded acres or acres planted for harvest." Then, for each county in the domain, a crop distribution was calculated which gave the percent of the total crop area in a county that was made up of each crop type. This percent crop contribution was then used in the final calculation of the emission rate for that county's crop area. The crop area found in the USDA NASS was not used. The total crop area planted for an individual county indicated by the USDA NASS did not equal the total crop area indicated by the Texas Parks and Wildlife Department vegetation data. This could not be resolved spatially so the percent contributions indicated by the USDA NASS were used for the area indicated by the TP&WD data. Table 2.14 shows the crop composition of the agricultural areas for each county.

Table 2.14: County crop compositions

County	County FIPs Code	Cover Codes	Percentage of County's Crop Area Planted for Specific Crop Type						
			Corn	Cotton	Oats	Peanut	Sorghum	Soy-Beans	Winter Wheat
Bosque	35	202	0.07	0.00	0.14	0.00	0.08	0.00	0.72
Clay	77	203	0.00	0.07	0.01	0.00	0.00	0.00	0.93
Collin	85	204	0.34	0.06	0.02	0.00	0.29	0.02	0.26
Cooke	97	205	0.07	0.00	0.12	0.00	0.33	0.06	0.42
Coryell	99	206	0.15	0.00	0.13	0.00	0.22	0.00	0.50
Dallas	113	207	0.29	0.05	0.02	0.00	0.24	0.07	0.34
Denton	121	208	0.11	0.00	0.00	0.05	0.44	0.02	0.38
Ellis	139	209	0.35	0.39	0.00	0.00	0.09	0.01	0.15
Erath	143	210	0.07	0.00	0.14	0.72	0.00	0.00	0.07
Fannin	147	211	0.15	0.07	0.00	0.08	0.19	0.27	0.24
Grayson	181	213	0.21	0.00	0.01	0.05	0.38	0.04	0.32
Hamilton	193	214	0.00	0.00	0.42	0.00	0.16	0.00	0.42
Hill	217	216	0.24	0.21	0.01	0.02	0.27	0.00	0.25
Hunt	231	218	0.07	0.41	0.00	0.00	0.20	0.06	0.26
Jack	237	219	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Johnson	251	220	0.14	0.08	0.04	0.00	0.35	0.00	0.39
Kaufman	257	221	0.13	0.34	0.02	0.00	0.16	0.15	0.21
Lampasas	281	222	0.00	0.00	0.45	0.00	0.00	0.00	0.55
Limestone	293	224	0.39	0.27	0.02	0.00	0.08	0.00	0.24
McLennan	309	225	0.44	0.11	0.07	0.00	0.13	0.03	0.22
Milam	331	226	0.30	0.40	0.00	0.00	0.22	0.00	0.07
Montague	337	227	0.00	0.00	0.00	0.13	0.00	0.00	0.87
Navarro	349	228	0.14	0.53	0.00	0.00	0.24	0.00	0.08
Parker	367	230	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Robertson	395	232	0.17	0.70	0.01	0.00	0.10	0.00	0.02
Rockwall	397	233	0.28	0.22	0.00	0.00	0.25	0.00	0.24
Tarrant	439	235	0.00	0.09	0.00	0.00	0.29	0.00	0.62
Wise	497	237	0.00	0.00	0.09	0.23	0.16	0.00	0.52

Table 2.15 gives the emission rates used for the crops from the USDA-NASS. The emission rates in Table 2.15 were obtained from data published in the literature (Lamb *et al.*, 1993). These emission rates were used to calculate

emission rates for the individual counties based on each counties crop composition.

Table 2.15: Emission rates given for crops in USDA-NASS data found in Texas counties (Lamb, 1993)

CROP	Biomass ($g\ m^{-2}$)	Emission rate ($\mu g\ g^{-1} h^{-1}$)	Composition of Emission (Percent)			
			Isoprene	α-Pinene	Other Monoterpenes	Other VOCs
Corn	1610	1.1	0	10	10	80
Soybeans	740	0.03	100	0	0	0
Wheat	740	0.041	30	50	10	10
Sorghum	3180	0.012	38	20	25	25
Barley	1290	0.029	20	25	25	30
Oats	750	0.051	20	25	25	30
Cotton	160	0.237	20	25	25	30

The crop category of the Texas Parks and Wildlife Department's vegetation coverage was improved using the USDA-NASS data. This was done by dividing the crop area of the coverage into separate polygons for each county. Each of these polygons represented a single county's crop make-up, and a three-digit land use code was given for each polygon. This three-digit land use code was used to reference the specific crop composition, and thus the specific emission rate, for each county. The land use codes (201-237) were randomly assigned to the 37 county crop areas. The codes are shown in Table A.1 and Table A.2 of Appendix A.

This part of the database construction was accomplished with the use of Arcview and Arc/Info. First the crop category was separated from the rest of the coverage using Arcview. The coverage was displayed in a view and the select feature icon was selected while the TP&W coverage theme was raised in the legend. The crop polygons turned yellow as they were selected. Next the **convert to shapefile** option under the theme menu was selected. This made the selected polygon into a separate shapefile.

ArcInfo was then used to divide this individual polygon into different polygons for each county. First, the shapefile was converted to a coverage using the **shapearc** command in Arc.

Arc: **shapearc crops.shp crops**

Then the coverage topology was built.

Arc: **build crops**

Next, a polygon coverage of the county lines was used to divide the crop category up into individual polygons. This was done with the **intersect** command.

Arc: **intersect crops cntylines newcrops**

This left the coverage newcrops which was a polygon coverage of only the areas shared by the two original coverages. Figure 2.7 shows the original and NASS modified crop area.

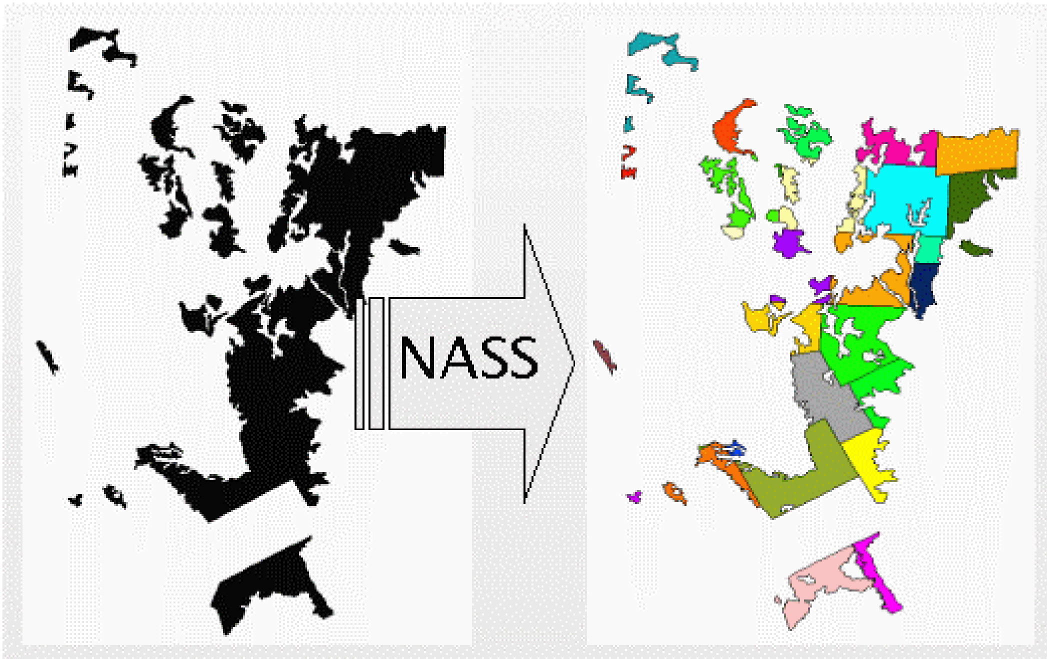


Figure 2.7: TP&WD crop region before and after modification using USDA NASS data.

The topology of the new coverage was built.

Arc: **build newcrops**

The new crop categories attribute table contained only the specific land use code for each county's agricultural area. These data were then ready to be attached back to the original Texas Parks and Wildlife Departments' vegetation coverage. This was done with the UNION command in Arc.

Arc: **union newcrops vegetation finalvegetation**

Again the coverage topology was built.

Arc: **build finalvegetation**

Figure 2.8 shows the TP&WD vegetation coverage with the modified crop category in gray. This TP&WD legend is given in Table 2.12.

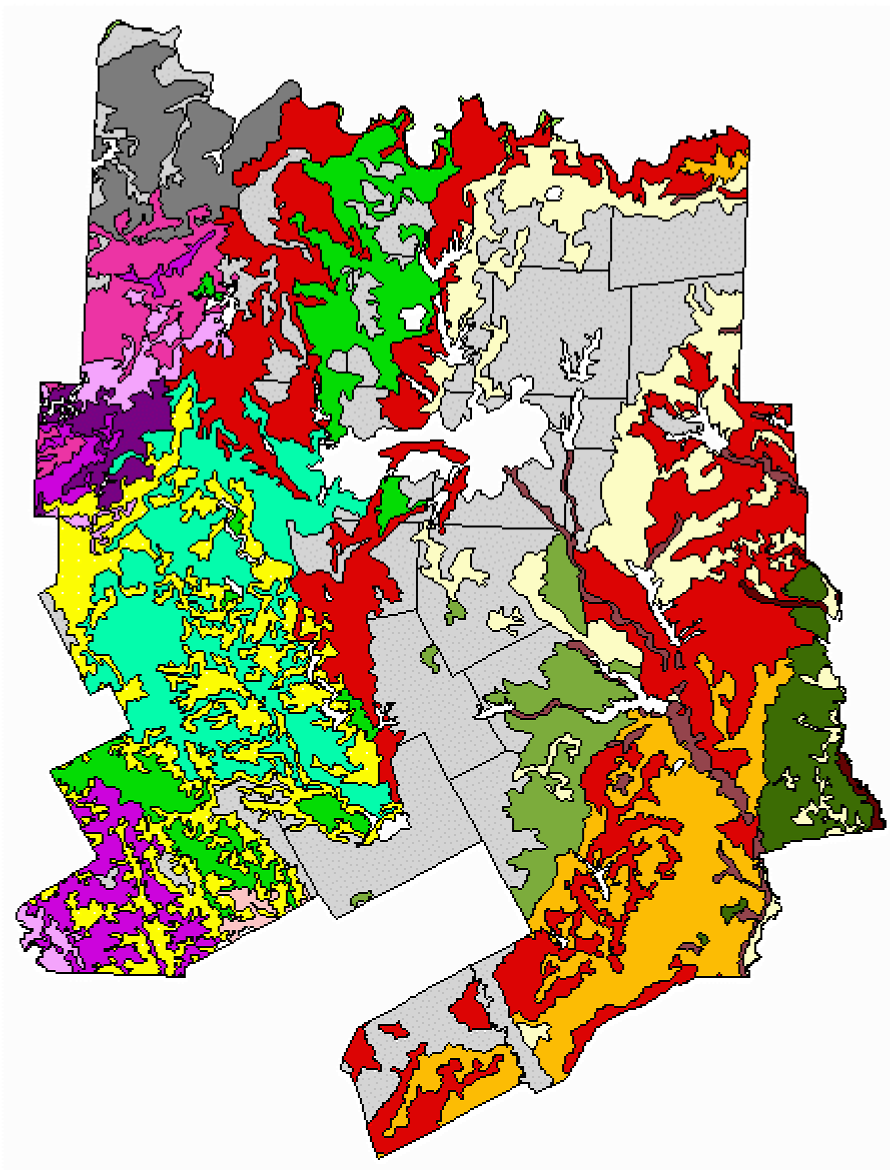


Figure 2.8: Texas Parks and Wildlife Coverage with modified crop region

2.6 CONSTRUCTION OF THE COMPOSITE DATABASE

The final coverage was completed by merging the NCTCOG land use coverage with the modified TP&WD vegetation coverage. The one major decision before this could be done was what to do with the areas classified as vacant in the NCTCOG land use coverage. It was decided that rather than leaving it in the final coverage and ground-truthing the vegetation for these areas, the TP&WD vegetation categories would be used in the areas where overlap occurred between the two coverages. To complete this task, both Arcview and Arc/Info were utilized.

First the TP&WD vegetation coverage was merged with the NCTCOG land use coverage to form one land use/ land cover coverage by using the **union** command in Arc.

Arc: **union nctcog finalvegetation final**

The topology of this new coverage was then built.

Arc: **build final**

The new coverage was taken into Arcview to allow easy modification of the attribute table. After the execution of the **union** command, the polygon attribute table gave both NCTCOG descriptions and TP&WD descriptions for every polygon in the coverage. A new field was created to have one field of the database giving data for both sets of attributes. This compilation field was to let the TP&WD categories describe the areas where the NCTCOG vacant land use was. This was done in Arcview using the **start editing** function under the edit menu. A new field was added to the attribute table using the **add field** command. The **query builder** was then used to assign values to this field such that the

NCTCOG assignments took precedence over the TP&WD assignments unless the NCTCOG vacant assignment was given. In this case the TP&WD assignments were given precedence. The NCTCOG vacant classification was eliminated because the vegetation classification of the TP&WD coverage was thought to be more valuable in the assignment of biomass, species distributions, and emission rates. The final biogenic emission estimation database for the North-Central Texas study domain is shown as Figure 3.1 in the results section of this document. Each polygon in this composite database was assigned one specific land use code that modelers can use to assign biomass and emission rate values. These data are shown in Table A.1 of appendix A.

2.7 Gridding the Study Domain for Input into the Emissions Models

Emissions models require an accurate spatial representation of the land use/ land cover categories as one of their inputs. This was accomplished by gridding the final domain and then changing the grid into an ASCII file that could be read by the emissions program. A FORTRAN program was written to transform the ASCII output representing the final coverage into a format which could be read by the emissions program. The land cover codes were used to interpret the characteristic land cover of the area throughout this process.

The domain was gridded into 0.5 km by 0.5 km cells for the coarse vegetation region. The Dallas/Ft. Worth urban region was gridded into 0.1 km by 0.1 km squares. It was believed that this gridding resolution did not lose significant amounts of the spatial resolution, and at the same time, it did not create tremendously taxing file sizes. The gridding of the coverage was accomplished by using the **polygrid** function in Arc.

Arc: **polygrid urban gurban 0.1 lu_code**

Arc: **polygrid rural grural 0.5 lu_code**

The gridded coverage did not maintain the properties of the original coverage. Only one attribute could be converted to describe each gridcell. The land use code was the attribute assigned to the gridcells. The last word in the above commands indicates this. The grids were then converted into ASCII files containing one attribute number in the place of each gridcell. These attribute numbers did not lose the spatial relationships. The ASCII files contained as many rows and columns as were in the original grids. A FORTRAN code written by ENVIRON was then used to convert the ASCII data into correct input format for the emissions estimation program.

Arc: **gridascii grural rural**

Arc: **gridascii gurban urban**

Once the ASCII input files were prepared via the FORTRAN program, another file containing the emission rates for individual land use codes was needed. This file was created based on biomass and species data collected in ground surveys of the domain. Emission rates were calculated from known emission rates published in the literature depending on the survey data. This emission file was also made available to the emission program, and the emission rates were accessed depending on the land cover codes of the database. The emission file and the required FORTRAN codes were developed by ENVIRON and are not presented in this document.

Another gridding technique could be used to prepare the data for input into the biogenic emissions model BIOME without the use of a FORTRAN code. This method consists of creating a grid in Arc/Info, intersecting the grid with the dataset coverage, and creating specific input tables from the intersected dataset.

A grid of any size, orientation, and having any cellsize can be created in Arc/Info using the **generate** command.

Arc: **generate < name>**

Generate: **fishnet**

Fishnet Origin Coordinate (X,Y): <__ , __>

Y-Axis Coordinate (X,Y): <__ , __> (Determines angle of grid)

Cell Size (Width, Height): <__ , __>

Number of Rows, Columns: <__ , __>

Generate: **quit**

Arc: **build <name>** (creates the grid)

Arc: **addxy <name>** (adds x and y coordinates of gridcell to the attribute table)

After the grid is constructed and its alignment is verified, it is merged with the land use coverage using the **intersect** command. This will add coordinate fields to the attribute table of the final coverage. BIOME input tables can be created from this attribute table using Arcview. Tables 2.16 and 2.17 show the format and contents of the required BIOME input files. These files should be in text format.

Table 2.16: Contents and format of Gridlu.dat BIOME input file

Variable Description	Variable Name	Format
Grid ID number	GRIDID	I4
X axis cell index	ICELL	I3
Y axis cell index	JCELL	I3
State FIPS code	STID	I2
County FIPS code	CYID	I3
Area of LUCODE in the grid cell (hectares)	LUAREA	R11
Land use code	LUCODE	C3
I = integer; R = real number; C = character		

Table 2.17: Contents and Format of Landuse.dat BIOME input file

Variable Description	Variable Name	Format
State FIPS Code	STID	I2
County FIPS Code	CYID	I3
Land use code	LUCODE	C3
X axis cell index	ICELL	I2
Y axis cell index	JCELL	I2
Area of LUCODE in county (square meters)	AREA	R11

3. RESULTS AND DISCUSSION

3.1 RESULTING FINAL COVERAGE

Figure 3.1 shows the final database constructed for the North Central Texas study domain. This database was the compilation of the Texas Parks and Wildlife Department vegetation data, the North Central Texas Council of Governments data, and the United States Department of Agriculture National Agriculture Statistics Service data. The North Central Texas Council of Governments data were used for the urban Dallas/ Fort Worth area. The Texas parks and Wildlife Department data modified with the USDA NASS data were used for the rural regions of the domain. Table 3.1 shows the land use/ land cover categories for this database (with all the counties' agricultural areas represented by gray for presentation). The actual 67 land use types used in the composite database are shown in Table A.1 of the appendices.

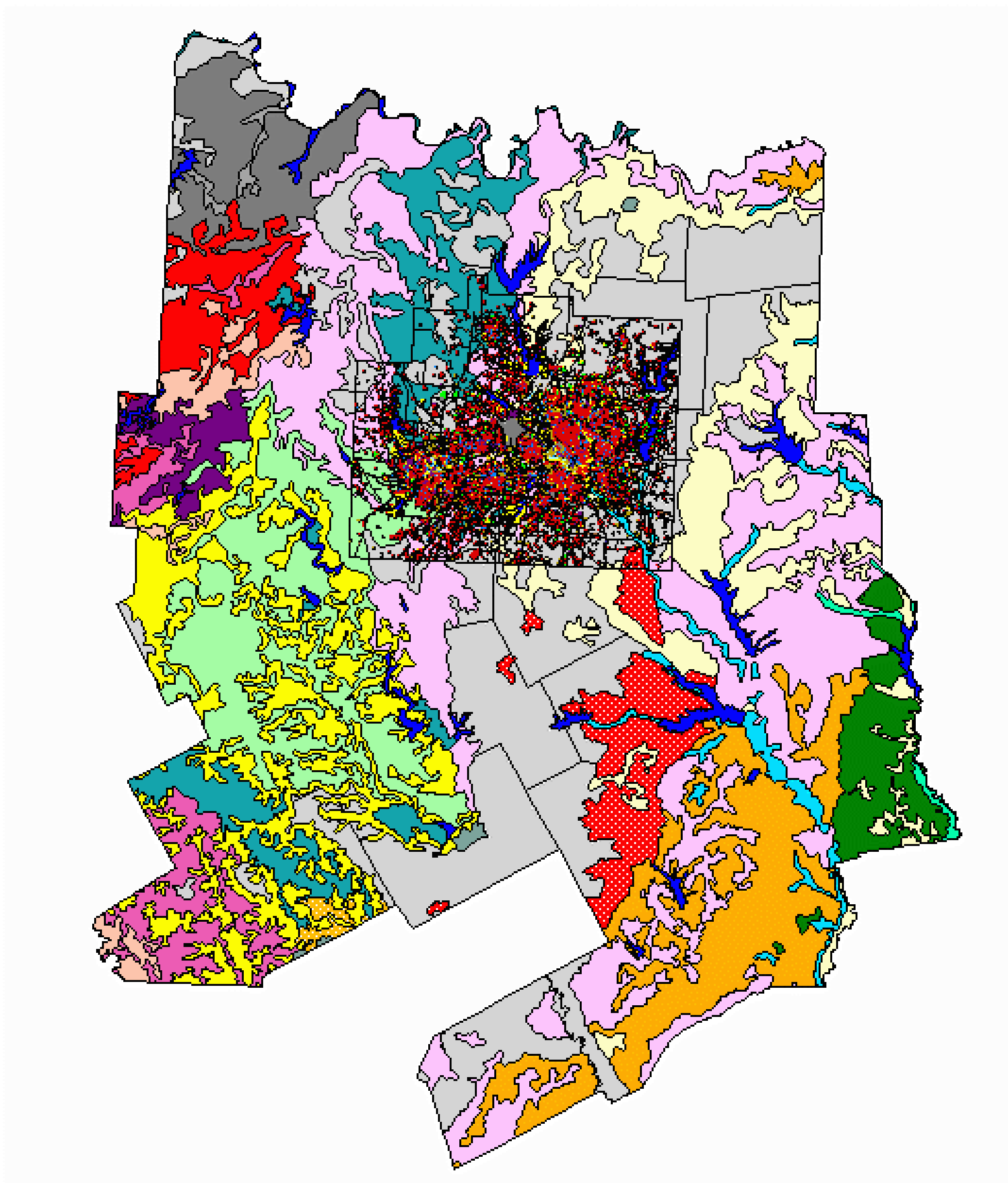
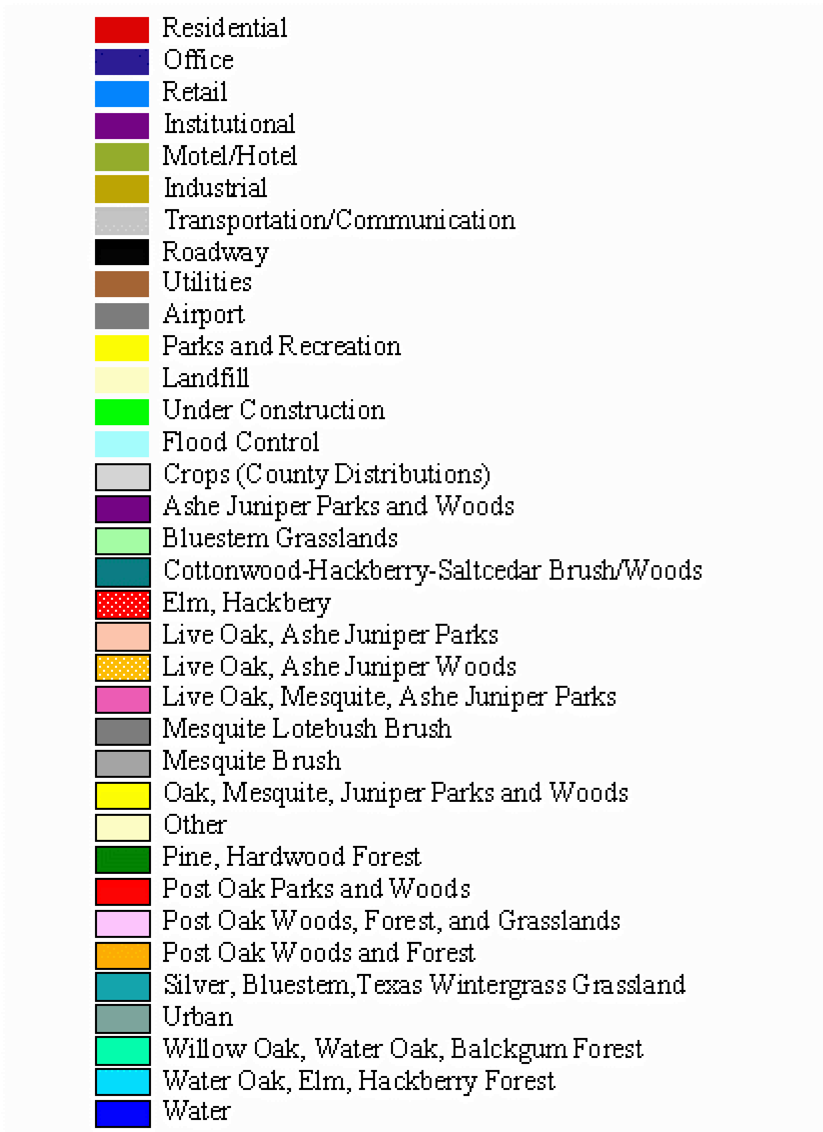


Figure 3.1: Land Use and Land Cover mapping of the biogenic hydrocarbon emissions database

Table 3.1: Final North-Central Texas Database Legend



3.2 COMPARISON WITH USGS LCC DATA

The USGS Land Cover Characteristics database was not used to describe the rural areas of the composite vegetation database for this study. Final assessment of the data showed that the Texas Parks and Wildlife Department data was more suitable for this project for several reasons.

The USGS LCC database was created as part of an effort to map the entire world's vegetation using satellite data. As of early 1998, no resources have been used to complete ground verification of the resulting data because of the tremendous scale of this project.

The resolution of the USGS LCC is one square km, a relatively high resolution for remote sensing vegetation data. There are over 200 classifications of vegetation given for the entire continent. However, the classifications for the study region were not diverse. Although over fifty USGS LCC classifications were cited in the study domain, only four classifications accounted for more than 83% of the total area. These are listed in Table 3.2. In contrast, it would take the 9 largest classifications of the TP&WD vegetation database to describe 83% of the study area. Table 3.2 illustrates this comparison.

Table 3.2: Comparison of dominant categories in TP&WD and USGS LCC

USGS LCC		**TP&WD**	
Classification	**Percent of Domain**	**Classification**	**Percent of Domain**
Grassland with Cropland	34.03	Crops (all counties)	22.57
Cropland (Corn, Soybeans, Cotton, Rice) with Woodlands	22.87	Post Oak Woods, Forest, and Grasslands	18.26
Grassland/Woodland (Oak) Mosaic with Cropland	16.51	Oak, Mesquite, Juniper, Parks and Woods	8.14
Cropland (Corn, Cotton, Sorghum, Pasture)/Grassland Mosaic	9.82	Other	7.78
		Bluestem Grasslands	7.43
		Post Oak Woods and Forest	7.27
		Silver, Bluestem, Texas Wintergrass Grassland	5.32
		Mesquite Lotebush Brush	3.22
		Elm and Hackberry	3.10
Amount of Domain	83.22 %	Amount of Domain	83.09 %

The comparison illustrated in Table 3.2 shows an incomplete description of the vegetation trends for the study domain in the USGS LCC. Furthermore, several categories have the same vegetation species identified. For example, each of the largest four vegetation categories in the USGS LCC include the cropland classification.

Another point of concern when considering the USGS LCC coverage is that although the vegetation assignments are in general, technically correct, the species of vegetation used in the descriptions are the herbivorous understory rather than the arboreal species. The crop and grassland vegetation categories, which dominate more than 83% of the coverage, identify the distribution of the lower-emitting species and these classifications do not help in identifying the locations of the more biogenically important trees and overstory. In contrast to this, the TP&WD vegetation focuses its vegetation classifications on the larger vegetation which has higher biomass and larger emissions.

Other researchers have used the USGS LCC vegetation data for their biogenic hydrocarbon emission estimates. However, the scope of their study was much greater than that of the North-Central Texas study. For example, Kinnee *et al.* (1997) used the USGS LCC database to establish their biogenic emissions estimates for the entire western half of the United States. The resolution of Kinnee's study was on the county level. This approach would not be as accurate for the Texas study because the vegetation for Texas changes rapidly over relatively small distances. Figure 3.2 shows a section of the study domain in which the vegetation varies between species with extremely different emission rates over a very short distance.

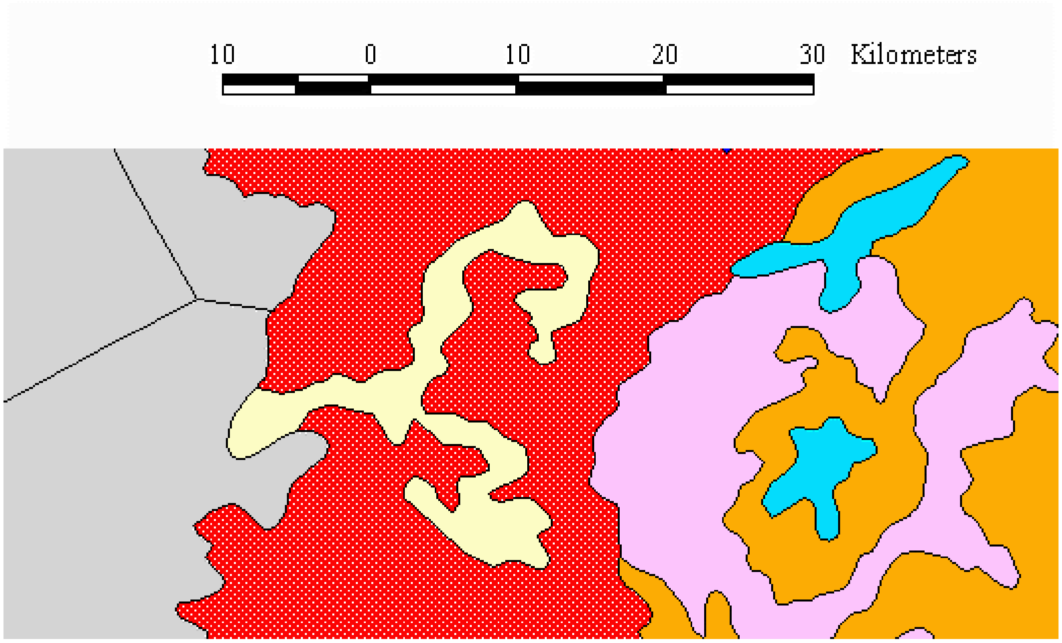


Figure 3.2: Vegetation in southeast section of Texas Parks and Wildlife Department data

Figure 3.2 shows that over a relatively small distance, the vegetation in the study domain can be quite different. Furthermore, as seen in Table 3.1, the red area in Figure 3.2 represents elm and hackberry, and the orange area represents post oak woods and forest. Both of these classifications have high biogenic hydrocarbon emission rates. The gray area represents crops. The emissions for the crops are much lower. Therefore, the most detailed spatial representation available is very important for accurate biogenic emission estimation.

3.3 COMPARISON WITH LANDSAT DATA

As part of the North-Central Texas biogenic emissions study, a Landsat satellite image of part of the region was purchased. Although the image was not useful in identifying specific vegetation species for the study, it was beneficial. The Texas Parks and Wildlife Department's vegetation coverage is derived from a combination of satellite data, geology, and ground surveying. Therefore, the TP&WD coverage was expected to have trends very similar to the Landsat data that were collected for the project.

The Landsat image was analyzed in two separate sections by researchers at the National Center for Atmospheric Research (NCAR). The rural area of the image was separated from the urban area. This was done because the concrete and metal surfaces found in the urban areas have a higher reflectance of the satellite wavelengths. This causes a brighter image that is not sensitive to subtle spectral variations. However, the less sensitive rural areas of the image can be enhanced by these spectral variations. The rural section of the purchased Landsat image is shown in Figure 3.3. The missing upper left-hand section of the rural image was masked by cloud-cover. The section missing in the center of the figure is the urban area and is shown in Figure 3.4. The color legends for the two sections of the Landsat image are given in Tables 3.3 and 3.4.

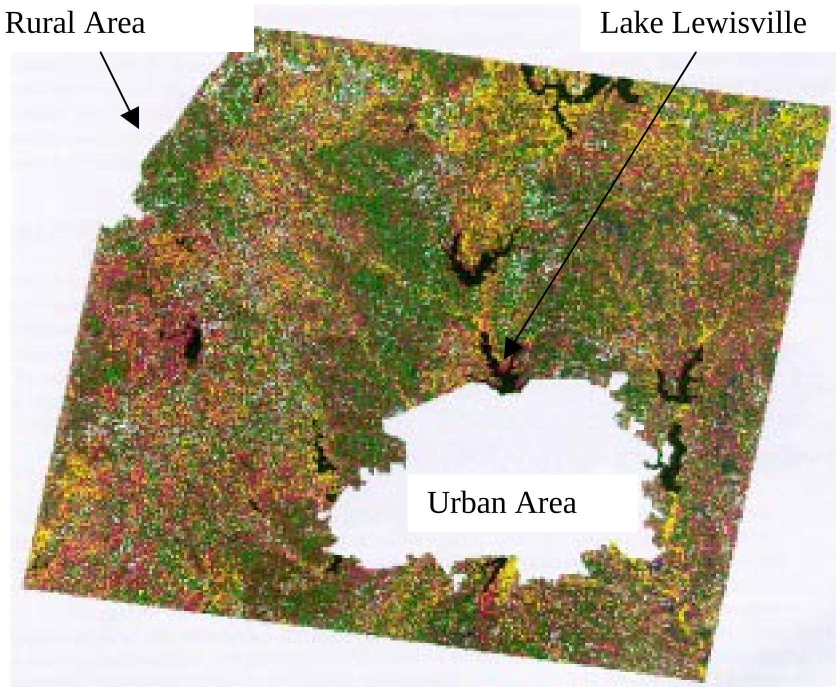


Figure 3.3: Rural section of Landsat image purchased for NCT study

Table 3.3: Rural area classifications of the Landsat TM data

Class	Percent Cover	Description	Pixel Color
0	N/A	Unclassified Background	Grey
2	3.1	Open water	Black
3	0.8	Damp bar soil; little vegetation (e.g. shorelines)	Royal Blue
4	4.9	Trees; high amount of green leaf biomass	Yellow
5	7.1	Trees; high amount of green leaf biomass (but slightly lower than 4)	Yellow
6	7.2	Parkland/Woodland, lower green tree leaf biomass than 5	Pink
7	11.6	Open parkland savanna, lower green tree leaf biomass than 6	Pink
8	6.7	Senesced/ cleared croplands	Forest Green
9	21.0	Non-crop, likely open range with low green biomass (e.g. rangeland)	Dark Brown
10	20.2	Sparsely vegetated (herb-dominated), high soil or senesced component	Light Brown
11	11.7	Intermediate green (herbaceous cropland) vegetation, greener than 10	Green
12	5.6	Recently cleared agricultural area (bare soil) and cement	White

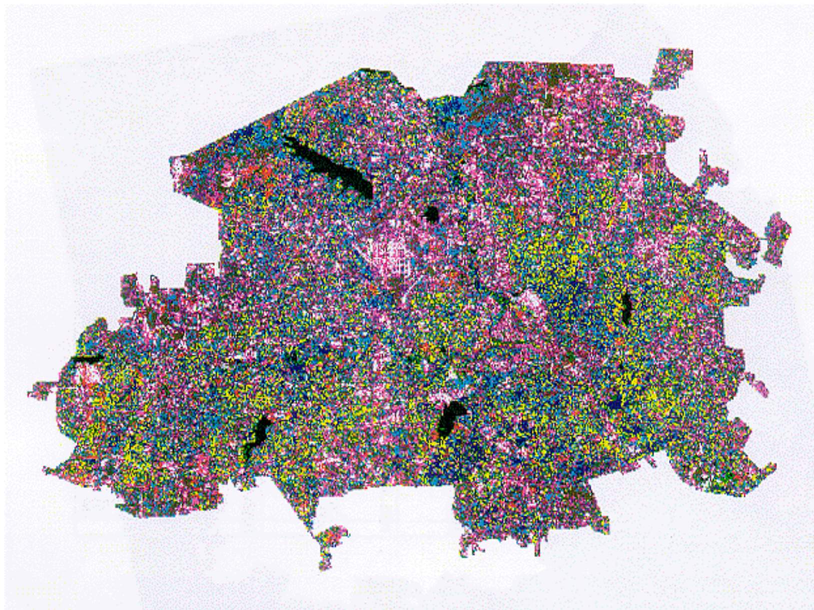


Figure 3.4: Urban section of Landsat image purchased for North Central Texas study

Table 3.4: Urban area classification of the Landsat TM data

Class	Percent Cover	Description	Pixel Color
0	N/A	Unclassified background	Grey
1	1.8	Open water	Black
2	1.6	Shallow/turbid water	Green
3	9.3	Green vegetation (likely trees; perhaps conifers), high soil moisture	Royal Blue
4	10.2	Green vegetation (likely deciduous trees)	Yellow
5	13.9	Less green (tree) vegetation than 4, higher bare soil component	Turquoise Blue
6	2.1	Barren/ open soil, cleared fields with little (herbaceous) green vegetation	Brown
7	8.5	Intermediate (likely herbaceous) green vegetation with small amount of concrete	Maroon
8	14.3	Maximum (likely herbaceous) green vegetation	Forest Green
9	12.5	Intermediate (likely herbaceous) green vegetation with small amount of concrete	Purple
10	4.8	Intermediate (lower than 9, likely herbaceous) green vegetation with more concrete	Orange
11	15.0	Low green vegetation (likely herbaceous) with mixture of concrete	Pink
12	6.0	Concrete with no vegetation component	White

Although the images were not able to identify specific species of vegetation in the domain, the Landsat imagery was useful in validating the use of the Texas Parks and Wildlife Department vegetation database. The corresponding section of the TP&WD data is shown in Figure 3.5.

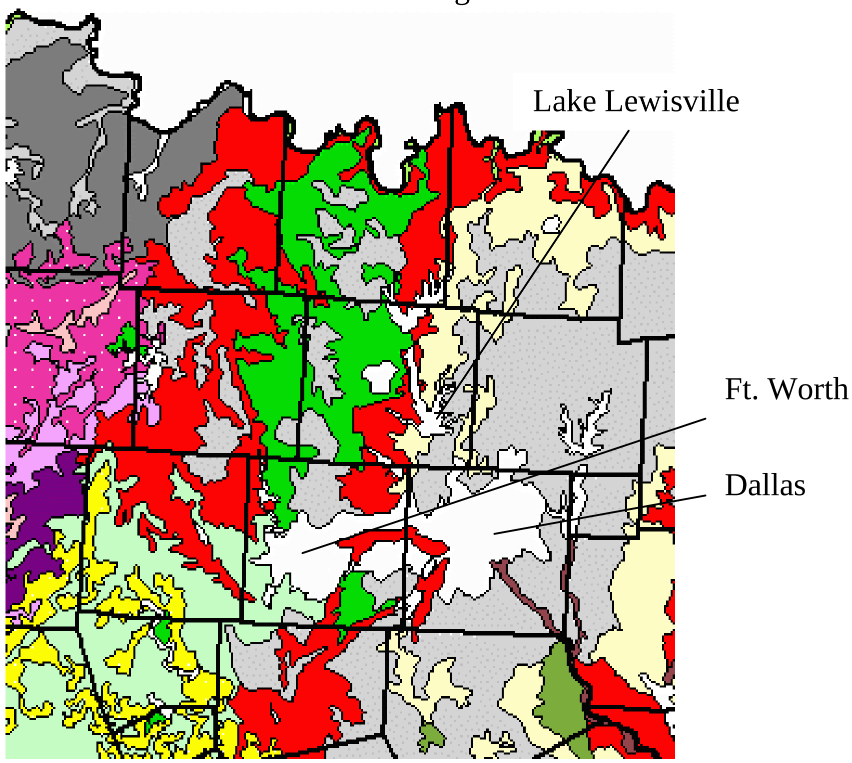


Figure 3.5: Area of Texas Parks and Wildlife Database that corresponds to purchased Landsat imagery

A qualitative comparison of the TP&WD data and the USDA NASS data was done. The lakes in the Landsat imagery are black, as is the rest of the water, while in the TP&WD data the lakes are white (other than the large white area representing the Dallas/ Fort Worth urban area). For a reference, Lake Lewisville is labeled in both views. Using the lakes as reference points, some general trends were observed in both data sets. The red area going north from the Metroplex and passing by Lake Lewisville and Ray Roberts Reservoir is post oak woods forest and grassland. The corresponding area in the rural Landsat image (Figure 3.3) is coded yellow with a small amount of pink. Table 3.3 shows that these colors indicate trees, woods and forest and other areas of high green leaf biomass.

Another qualitative comparison can be made by looking at the area just to the left of this region. The TP&WD vegetation data shows a green area with gray interspersed running northward from Fort Worth. The gray represents cropland, and the green represents Silver, Bluestem, and Texas Wintergrass grassland. The corresponding area on the rural Landsat imagery is color-coded green and brown (Figure 3.3). Table 3.3 shows that those colors represent herbaceous cropland, or sparsely vegetated soil with low green leaf biomass.

Both of these qualitative comparisons show that the Landsat imagery generally supports the vegetation distribution for the TP&WD data. Although this should be expected because the TP&WD data were derived from Landsat data, these trends are encouraging because the satellite data for this project was taken 15 years later than the original Landsat images from which the TP&WD data were derived.

3.4 GROUND SURVEY VERIFICATION

The satellite data demonstrated the accurate distribution of the vegetation, however there is no way to verify vegetation species with one Landsat image. After the vegetation database was constructed for the study domain, a ground crew surveyed the individual vegetation categories as part of the study. Although this document will not discuss the details of the biomass or emission rate assignments, the species classifications helped to emphasize the differences between the TP&WD vegetation categories.

Table 3.5 shows the species found in two ground surveys done in the western part of the domain. The TP&WD vegetation categories represented in this table are the post oak woods and forests and the ashe juniper parks and woods. As can be seen in the composite database shown in Figure 3.1, the classifications are adjacent to each other on the western side of the database (represented by the purple and red color coding respectively).

Table 3.5: Vegetation identified in separate categories of the Texas Parks and Wildlife Department coverage

Scientific and common names found in post oak woods and forests	Scientific and common names found in ashe juniper parks and woods
Vaccinium arboreum (Farkleberry)	*Prosopis glandulosa* (Honey Mesquite)
Quercus falcata (Southern Red Oak)	*Juniperus ashei* (Ashe Juniper)
Carya texana (Black Hickory)	
Bumelia lanuginosa (Woollybucket Bumelia)	
Quercus nigra (Water Oak)	
Ilex vomitoria (Yaupon Holly)	
Quercus marilandica (Blackjack Oak)	
Fraxinus americana (White Ash)	
Ulmus alata (Winged Elm)	
Quercus stellata (Post Oak)	
Juniperous virginiana (Eastern Redceder)	

The species identified in Table 3.5 were determined to be representative of the area vegetation category. This table shows that the species found in these two adjacent vegetation categories are very different from each other and that these areas require the separate vegetation classifications given to them in the Texas Parks and Wildlife Department data.

Overall, the ground surveys found several distinct differences between the vegetation species in the different vegetation categories. There were also significant differences between the classifications of the North Central Texas

Council of Governments database. Table 3.6 shows the different amounts of foliar biomass that were found in the Dallas/ Ft. Worth area. These data emphasize the varying amount of vegetation biomass in these land use categories which is also important in the calculation of biogenic emissions.

Table 3.6: Biomass densities assigned to urban land use classifications

Urban Land Use Category and Anderson Level Three Land Use Code		***Total biomass density (g/m2)***
Institutional	123	11.3
Commercial	122	14.9
Industrial	131	3.2
Parks and Recreational	171	78.3
Residential	110	42.2

3.5 Initial Emission Estimates

This section presents the initial emission grids found with the biogenic emission models using the database developed in this study. The gridded database was applied to the emissions models via a FORTRAN program written by ENVIRON, a partner contractor in this work. The biomass and species characterizations determined by the field team were also given to ENVIRON. Researchers from ENVIRON and the National Center for Atmospheric Research then established emission rates for each category in the final database based on the field data and emission rates published in the literature. The emission rates for each land use category were submitted to the emissions model in a file accessible to the emissions program via the FORTRAN program.

An old emissions grid for part of the study domain is shown in Figure 3.6. This emissions plot was generated using the BIES 2 biogenic emissions model, and has a county level resolution for its vegetation data. The BEIS 2 model uses the BELD land cover data as its vegetation characterization.

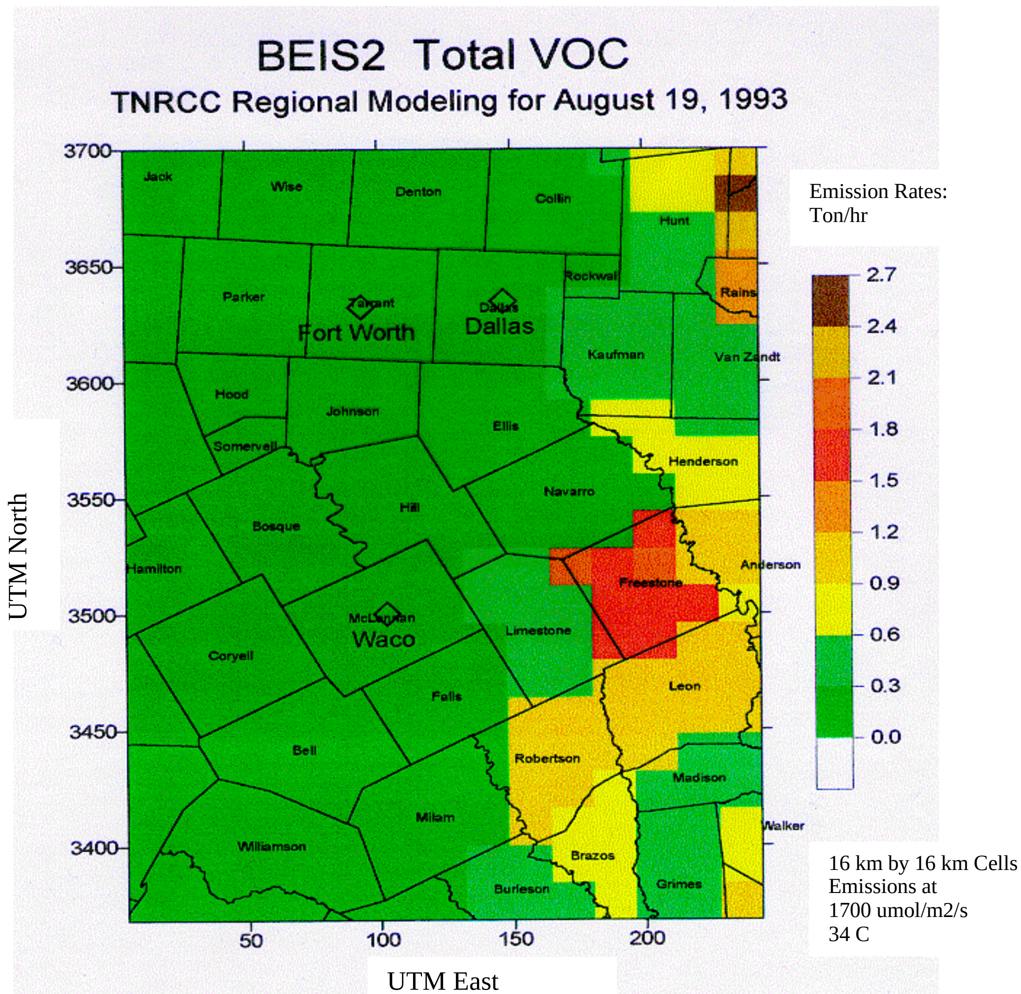


Figure 3.6: BEIS 2.0 biogenic emissions grid for part of the North-Central Texas domain

The scale shown in Figure 3.6 only shows relative emission rates and is not quantitatively correct. The BELD vegetation database considers the western half of the United States to be arid. The vegetation is not well defined in the western half of Texas. This trend can be seen in Figure 3.6. The emissions are low for the west side of the study domain. This is due to the incomplete state of the vegetation database rather than the lack of emissions in this area. Also, because 16 km square grid cells are used for BEIS 2, the emissions are averaged over a large area.

Figure 3.7 shows the preliminary emissions grid using the data developed for this study. Both the emissions and the resolution are different for the new vegetation database.

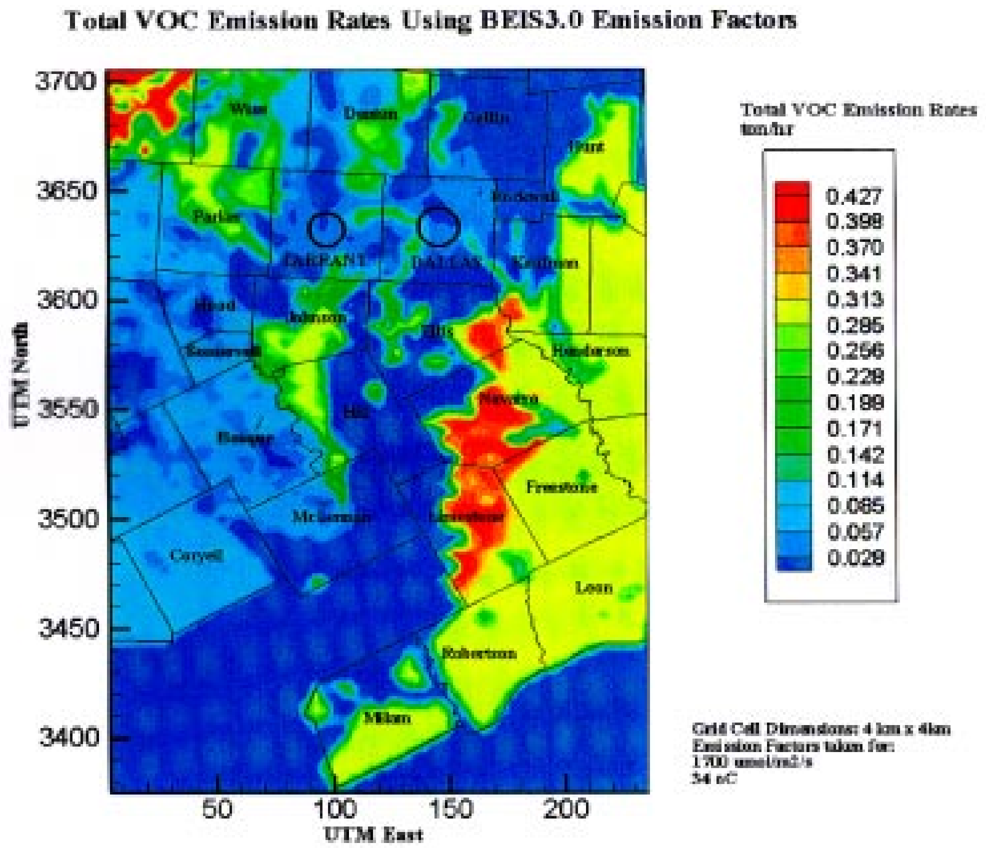


Figure 3.7: Emissions grid developed with the new North-Central Texas vegetation database

Figure 3.7's preliminary emissions estimates were calculated using 4 km grid cells. The resulting emissions plot shows that this higher resolution was needed to accurately estimate the biogenic emissions for the study domain. Figure 3.7 also shows that there is a relatively high emitting area in the northwest corner of the domain. This was a result of the large volume of high emitting post oaks in that area. The area was previously neglected by the BEIS 2.0 model and demonstrates the need for the improved vegetation classification for this region.

4. APPLICATIONS

4.1 APPLYING THE FINAL COVERAGE

The primary application of the vegetation database produced in a GIS in this thesis was to give an estimate of biogenic hydrocarbon emissions for the 37 North-Central Texas domain. As mentioned before, once the final vegetation database was constructed, each land use or land cover category was assigned both a foliar biomass and emission rate. Fieldwork was done to characterize the vegetation and calculate the biomass of each land use category. The foliar biomass was calculated using diameter at breast height algorithms developed in the literature (Geron *et al*., 1994). The vegetation characterization was used to apply emission rates to the land use categories taking into account the species composition of each category and the emission rates found in literature for the species (Benjamin *et al*. 1996). The final emissions were then calculated using one of several models which took into account the temperature, leaf area index (LAI), photosynthetically active radiation (PAR), biomass of the category and the biomass and chemical emission rates of each land use land cover category. The preliminary biogenic hydrocarbon emission estimates are shown in Figure 3.7 in chapter 3.

Several additional GIS applications have been utilized over the course of this project. One useful GIS application is that spatial data taken from a global positioning satellite system can be easily utilized. When vegetation surveys were conducted for biomass and species classification of the final database categories, GPS readings were taken to verify the location of the surveys. The coordinates of these readings could be used to easily identify the location of the survey within the GIS database. The positional accuracy of the Garmin GPS 12XL unit used was 15 meters (49 ft) with a possible degradation to 100 meters. The layout of the rural surveys is shown in Figure 4.1 where the white circles with the black dots represent the survey locations. The southern-most survey was taken outside of the domain in an accessible part of the post oak woods, forest and grassland.

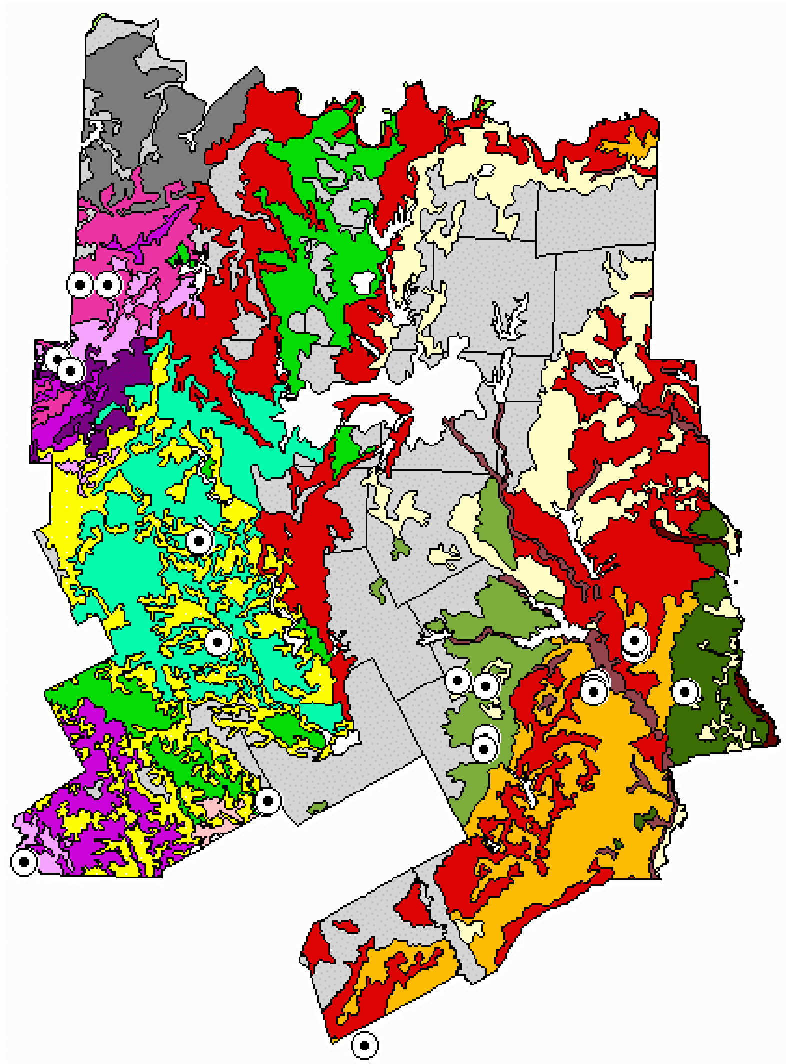


Figure 4.1: Locations of North-Central Texas Ground Surveys.

Another potential application of the type of work done for this research is the revision of land use files for an urban airshed model. An urban airshed model (or UAM) is an EPA approved model which gives the most accurate air quality modeling for a domain. The model is designed to take into account the biogenic emissions, anthropogenic emissions, and the climatology of an area and perform photochemical modeling to calculate the temporally and spatially varying quality of the air, specifically the ozone concentrations. Table 4.1 and 4.2 describe the include, input, and output files used in the execution of the UAM Biogenic Emission Inventory System version 2 (BEIS-2).

Table 4.1: Include Files Required for UAM BEIS-2

Name of File	Description
GRIDDEF.INC	Contains parameters used to define modeling grid. Grid is in rows and columns. Subsections of domain can be specified.
PARAM.INC	Parameter set related to the land use and emission factor files. Number of vegetation types and number of emission classes are defined, as well as array element positions for certain pollutants.
CRPARAM.INC	Indicates number of canopy layers and number of tree types for land use file. Array indexes for major land use types are set as well.
TIMESTEP.INC	Start date and time step are placed in a common block in this file.
SWITCH.INC	File describes a common block for the SWITCH array.

Table 4.2: Input and Output files for UAM BEIS-2

File Name	Input or Output	Description
FROSTFL	Input	ASCII formatted file stating first and last frost dates for year (required only for winter emissions).
LANDUSE	Input	Detailed ASCII formatted file containing the number of hectares in each grid cell for each land use category in that cell.
SUMFAC	Input	File contains summer emission factors for each land use type in LANDUSE. The emission rates are for Isoprene, Monoterpene, VOC, and NO.
WINFAC	Input	File with winter emission rates in same format as SUMFAC. (required only for winter emissions)
SURTEMP	Input	ASCII formatted file giving hourly, gridded surface temperatures in Kelvin.
SOLRAD	Input	ASCII formatted file giving hourly, gridded, solar radiation.
BIOEMIS	Output	Output file containing gridded biogenic emissions data adjusted for temperature and radiation in moles-C/hour.
NORMEMFL	Both	ASCII formatted file that contains emissions data not corrected for temperature or solar radiation. It is an output of the UAM BIES-2 program, but can also be used repeatedly as an input for UAM BEIS-2 for different meteorological scenarios.

The biogenic emissions estimated by a UAM are done with BEIS-2.0. The BEIS model's default uses the BELD data, which gives countywide vegetation cover with county wide emission rates to obtain emission estimates for an area similar to those shown in Figure 3.6. The LANDUSE file and emission files give the user the opportunity to update the land cover with more detailed data, and the SUMFAC and WINFAC files give the user the opportunity to update the emission rates for the land use categories. The UAM BEIS-2.0 user's manual states that spatial allocation software is required to create the gridded land use files.

4.2 CENTRAL TEXAS BIOGENIC EMISSION ESTIMATION DATABASE

Once the database was completed for the North-Central Texas modeling domain, the focus of the study was shifted to the south. Austin and San Antonio are also Texas cities which are near non-attainment for the National Ambient Air Quality Standard for Ozone. In an effort to better understand the air quality of central Texas, a biogenic emission database has been constructed for the Alamo urban airshed modeling domain (see Figure 4.2).

The methodologies used to construct a biogenic emission database for central Texas were the same as those used for North-Central Texas. The four databases used are listed in Table 4.3.

Table 4.3: Data sources used in the construction of the central Texas biogenic emissions database.

Data	Source	Scale	Last Updated
Texas Parks and Wildlife Department Vegetation Coverage	TP&WD http://www.tnris.state.tx.us/gispage.html	1:250,000	1984
USDA-National Agricultural Statistics Service (NASS)	USDA http://www.usda.gov/nass/nassinfo/nassinfo.htm	County	1996
USGS Land Use Land Cover	USGS http://edcwww.cr.usgs.gov/doc/edchome/ndedb.html	1:250,000	1974-1978
Austin Land Use Land Cover	City of Austin	Derived from 30 meter data	1990

The difference between the central and north central Texas vegetation databases is the data used for the urban areas. The data sources for urban data in the central Texas database were the City of Austin and the USGS LULC for the Austin and San Antonio areas respectively. Although the USGS LULC data was completed in the 1970's, it was used for the San Antonio area because at the time of the research an updated land use/land cover data source could not be found for San Antonio. This part of the domain can be updated as soon as more recent data is identified.

The construction of the Central-Texas vegetation database was done using the same methodology as the North-Central Texas database. The Texas Parks and Wildlife Departments vegetation database was determined to be the superior rural vegetation data set during the North-Central Texas work and was used for this database as well. The crop category of the TP&WD database was similarly divided into the county specific regions. 1997 crop data from the USDA-NASS were used to obtain the most current crop distributions possible. Finally, the Land Use data for the Austin and San Antonio domains were added into the final Central-Texas vegetation database. Figure 4.2 shows the biogenic hydrocarbon emissions database constructed for the Austin-San Antonio modeling domain.

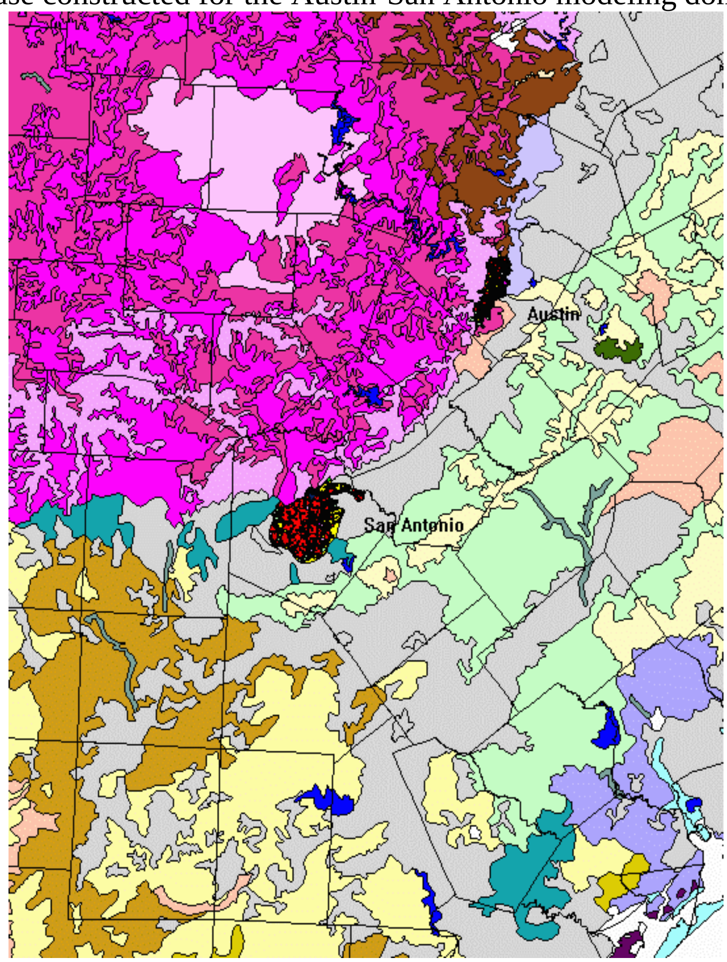


Figure 4.2: Composite Austin-San Antonio biogenic emissions database

Table 4.4: Austin-San Antonio database legend

Blue Gramma and Buffalo Grassland
Bluestem Grassland
Ceniza, Blackbrush, Creosotebush Brush
Crops
Elm, Hackberry Forest
Live Oak Woods/Parks
Live Oak, Ashe Juniper Parks
Live Oak, Ashe Juniper Woods
Live Oak, Mesquite, Parks
Live Oak, Mesquite, Ashe Juniper Parks
Marsh Barrier
Mesquite Blackbrush Brush
Mesquite, Granjeno Parks
Mesquite, Granjeno Woods
Mesquite, Juniper, Live Oak Brush
Mesquite, Live Oak, Bluewood Parks
Mesquite, Juniper Shrub
Oak, Mesquite, Juniper Parks/Woods
Other
Pecan, Elm Forest
Pine, Hardwood Forest
Post Oak Woods/Forest/Grasslands
Post Oak Woods/Forest
Silver Bluestem, Texas Wintergrass Grasslands
Urban
Water
Water Oak, Elm, Hackberry Forest

The composite Austin-San Antonio biogenic emissions database is derived from the TP&WD data, Austin LULC data, San Antonio area USGS LULC data, and the USDA-NASS data.

4.3 DISCUSSION

The vegetation species distribution for a region depends on several factors. Many of these factors, such as precipitation, elevation, and soil type, have spatial properties and can be represented in a GIS. Other factors such as temperature and photosynthetically active radiation can also affect the rate of biogenic hydrocarbon emissions.

The amount of precipitation in an area can have a large effect on the viability of a vegetation species growing in that area. Annual average rainfall data is usually readily available for a region. Figure 4.2 is an example of such a mapping for the state of Texas.

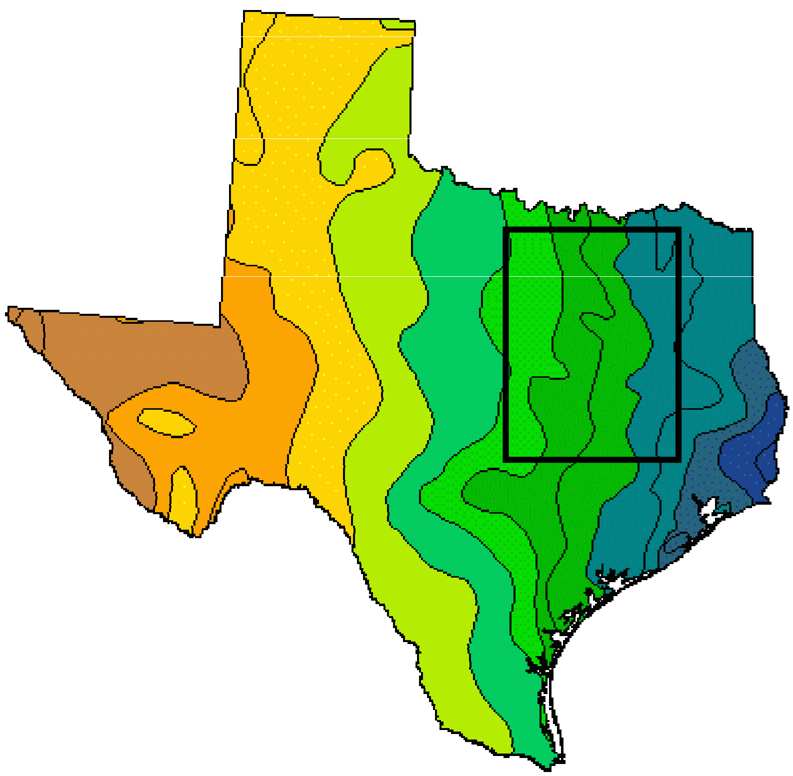


Figure 4.3: Precipitation for the state of Texas with approximate Dallas/ Fort Worth UAM modeling domain

Figure 4.3 shows the general precipitation trend for the state of Texas such that the inches of rainfall per year tend to decrease while traversing westward. The highest precipitation rates are in the blue regions of the map. The species of vegetation in the western part of the state are adept at growing under more arid conditions. However, across the North Central Texas study domain the annual precipitation rate does not appear to vary enough over the study region to be a significant factor in the distribution of vegetation.

The elevation can also be a factor in determining the species of vegetation that grows in a region. A GIS can represent changes in elevation with the use of digital elevation model data. However, changes in elevation are generally not great in Texas. The entire study domain is relatively flat with no mountains or other elevation changes large enough to effect the bio-diversity of the plant life.

The soil type of an area can also be a factor in determining the type of plant species that grows in a particular region. There is a large amount of soil data that is available for the entire US in GIS format. A comprehensive soil database for the entire United States is available from the USDA Natural Resources Conservation Service. The soil data is available in GIS format and is accompanied by several data tables. A GIS mapping of this data for the North-Central Texas study region is shown in Figure 4.4. These data could be useful in determining species distributions if accurate correlations could be developed. However, no such correlations currently exist.

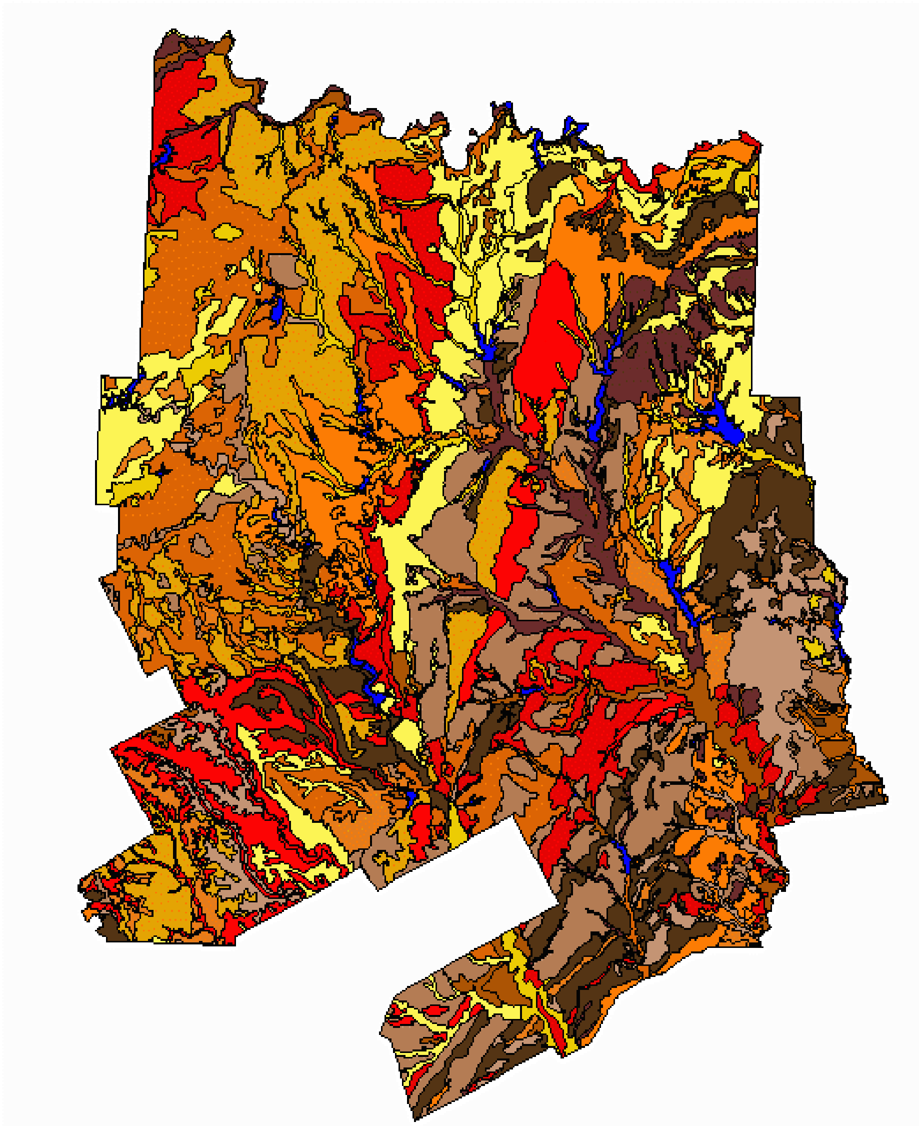


Figure 4.4: Texas STATSGO soils data for study domain. Polygons are delineated by map unit ID as specified by USDA NRCS digitizing standards.

The map units used for the image in Figure 4.4 need to be combined with other data tables to describe the soil represented by each polygon. One such table supplied with the STATSGO data lists some of the vegetation species found in the map units, but the vegetation data is very incomplete and often does not give an entry for every type of map unit. Nonetheless, the resolution of this data is at least as high as the resolution of the TP&WD data. Should any correlations between these two data sets be developed, the soil data might prove to be a useful tool in the development of vegetation distribution data for less characterized areas.

Currently, the emission models use algorithms from various researchers to take into account any changes in temperature or light intensity (PAR). The values of both these variables could be stored in a GIS. Although these variables fluctuate relatively quickly, the major trends for these data could be represented. Current biogenic emission models assume a constant value for each of these variables across the entire domain and calculate the estimates based on a peak temperature or PAR episode. A GIS could be used to take into account non-uniform parameters.

5. CONCLUSIONS AND RECOMENDATIONS

A geographic information system is a powerful tool that can be used as a database to store large amounts of information for each polygon, line, and point depicted in a coverage. It also is a powerful tool because of the spatial relationships that it maintains. This document has shown GIS applications useful in estimating biogenic hydrocarbon emissions. Future projects might explore the use of other GIS applications for representing vegetation distributions and for obtaining biogenic emission inventories.

Of the vegetation species distribution data now available for North Central Texas, as well as the rest of the state, the Texas Parks and Wildlife Department vegetation data proved to be the most useful. The data emphasized the emitting vegetation species, and the resolution of the data was desirable. Furthermore, survey work and comparisons with Landsat data showed that the TP&WD data had a very accurate vegetation distribution. Although the USGS land cover characteristics data had a much larger scope, the vegetation categories emphasized herbivorous vegetation. Also, the area of the study was dominated by a relatively few categories in the USGS LCC despite the large number of categories available.

This document discusses the methodologies used in developing a biogenic hydrocarbon emissions database. The methodologies are thought to be innovative and applicable anywhere in the state of Texas. The compilation of separate data sets into a final biogenic hydrocarbon emission estimation database was completed, and the spatial database was an integral part of the final biogenic emission estimations.

Once the GIS database was constructed in Arc/Info, it was possible to transfer this data into a biogenic emission model to produce a more detailed and accurate estimation of the total VOC inventory for the study region. This work proceeded smoothly due to the gridding and conversion features of Arc/Info.

Preliminary biogenic hydrocarbon emission estimations show relatively high concentrations of BVOCs being emitted southeast of the Dallas/Fort Worth metroplex. This is generally upwind of the metroplex, therefore these emissions play a role in the atmospheric photochemistry for the area. The vegetation distribution data developed in this study is an improvement over the vegetation distribution data currently being used in the UAM for the region, and urban airshed modeling for the region should be redone with the new data to evaluate the implications.

Future work should also be concentrated in the development of a reliable vegetation data set outside the state of Texas. The superior data set for this study was the Texas Parks and Wildlife study, however no data similar to this are available outside the state of Texas. The continent-wide data of the USGS land cover characteristics database provide detailed vegetation information, but the emphasis of this data is not on the arboreal, high emitting vegetation. Construction of databases for regions outside of the state of Texas will need to investigate data for the rural vegetation characterization other than the TP&WD data. Currently, the USGS Land Cover Characteristics data are preferred for vegetation species distribution in the western half of the United States. However, this study found that the USGS LCC does not provide enough information for accurate biogenic emission estimation. Western regions, and possibly all regions, outside of the state of Texas need a vegetation database as accurate and detailed as the TP&WD data. This development could be done by conducting large-scale ground truthing of the USGS LCC or by incorporating STASGO soil data into the vegetation analysis. The satellite data used in the construction of this data set could also be reprocessed to emphasize the larger species of plants.

Regardless of the method used to improve the vegetation species distribution data, the county wide vegetation data being used to estimate biogenic missions need to be improved. This study showed that biogenic emissions can vary over short distances due to changes in vegetation species distributions. It is also important for vegetation distribution data to be improved because they are crucial inputs for urban airshed models. The meteorological parameters and overall gridsize of the UAM demand a higher resolution of vegetation data than what the BEIS 2.0 default is currently providing. Although the methodology for combining advanced vegetation distribution data and urban airshed models is not straightforward, the resulting database allows a more accurate estimation of biogenic emissions. Thus, the effort should be made to upgrade these models because non-attainment areas are required to use the UAM as part of their state implementation plans.

APPENDIX A

Table A.1: Final land cover codes for composite North Central Texas database.

	Classification	*Description*	*Coverage Code*
FOREST	Ash_jun_pawo	ash juniper parks, and woods	401
	Bluestem_grass	Bluestem grass	402
	Cowood_berry_sc	Cottonwood, hackberry and salt cedar brush	403
	elm_hberry	elm and hackberry forests	406
	loak_ajun_pa	live oak and ash juniper parks	408
	loak_ajun_wo	live oak and ash juniper woods	409
	loak_mq_ajun_pa	live oak, mesquite, and ash juniper parks	411
	Mesq_loteb_sh	Mesquite and lotebush shurb	413
	Mesquite_brush	Mesquite brush	414
	oak_mq_jun_pawo	oak, mesquite, and juniper parks and woods	415
	Other	Other	416
	pine_hardwood	pine and hardwood forests	418
	Poak_pawo	post oak parks and woods	419
	Poak_wo_for_gr	post oak woods, forests, and grasslands	420
	post_oak_wofor	post oak woods and forests	421
	Silver_wintergr	silver, bluestem , and Texas Wintergrass grassland	422
	Urban	urban other than DFW	111
	Wiloak_waoak_gu	Willow oak, water oak and blackgum forest	424
	Woak_elm_hberry	water oak, elm, and hackberry forest	425
	Water	Water	500
URBAN	Single family	Residential	111
	Multi-family	Residential	112
	Mobile Home Parks	Residential	113
	Group Quarters	Residential	114
	Office	Corporate and government offices, banks	121
	Retail	retail trade and services: dept. stores,	122

		restaurants, etc.	
	Institutional	Churches, government facilities, museums, hospitals, schools, etc.	123
	Hotel/Motel	Hotels and motels	124
	Industrial	Manufacturing plants, warehouses, office showrooms, etc.	131
	Trans./Comm.	Railroads, radio and television stations, truck terminals	141
	Roadway	Includes all roads	142
	Utilities	Sewage treatment and power plants, water plants and systems, etc.	143
	Airport	Includes terminal and runways	144
	Parking Garages	Structures for storage of autos	145
	Parks & Recreation	Includes all public and private parks, golf courses, amusement parks, etc.	171
	Landfill	Sanitary landfills, land applications, waste management facilities	172
	Under Construction	land under construction	173
	Flood Control	Includes levies and flood channels	181
	Vacant	Undeveloped land	300
CROPS	Bosque	Bosque county crop distribution	202
	Clay	Clay county crop distribution	203
	Collin	Collin county crop distribution	204
	Cooke	Cooke county crop distribution	205
	Coryell	Coryell county crop distribution	206
	Dallas	Dallas county crop distribution	207
	Denton	Denton county crop distribution	208
	Ellis	Ellis county crop distribution	209
	Erath	Erath county crop distribution	210
	Fannin	Fannin county crop distribution	211
	Grayson	Grayson county crop distribution	213
	Hamilton	Hamilton county crop distribution	214
	Hill	Hill county crop distribution	216
	Hunt	Hunt county crop distribution	218
	Jack	Jack county crop distribution	219
	Johnson	Johnson county crop distribution	220
	Kaufman	Kaufman county crop distribution	221
	Lampasas	Lampasas county crop distribution	222
	Limestone	Limestone county crop distribution	224
	McLennan	McLennan county crop distribution	225
	Milam	Milam county crop distribution	226

	Montague	Montague county crop distribution	227
	Navarro	Navarro county crop distribution	228
	Parker	Parker county crop distribution	230
	Robertson	Robertson county crop distribution	232
	Rockwall	Rockwall county crop distribution	233
	Tarrant	Tarrant county crop distribution	235
	Wise	Wise county crop distribution	237

Table A.2: Final Database's Composition

Land Cover Code	**Classification**	**Percent of Study Domain's Area**
202	Bosque Crop Distribution	0.06
203	Clay Crop Distribution	0.59
204	Collin Crop Distribution	2.4
205	Cooke Crop Distribution	0.56
206	Coryell Crop Distribution	0.52
207	Dallas Crop Distribution	1.27
208	Denton Crop Distribution	0.74
209	Ellis Crop Distribution	1.85
210	Erath Crop Distribution	0.1
211	Fannin Crop Distribution	1.48
213	Grayson Crop Distribution	0.91
214	Hamilton Crop Distribution	0.01
216	Hill Crop Distribution	1.87
218	Hunt Crop Distribution	1.1
219	Jack Crop Distribution	0.09
220	Johnson Crop Distribution	0.81
221	Kaufman Crop Distribution	0.52
222	Lampasas Crop Distribution	0.04

224	Limestone Crop Distribution	0.89
225	McLennan Crop Distribution	2.23
226	Milam Crop Distribution	1.31
227	Montague Crop Distribution	0.44
228	Navarro Crop Distribution	0.73
230	Parker Crop Distribution	0.1
232	Robertson Crop Distribution	0.45
233	Rockwall Crop Distribution	0.33
235	Tarrant Crop Distribution	0.5
237	Wise Crop Distribution	0.67
401	ash_jun_pawo	1.13
402	Bluestem_grass	7.43
403	cowood_berry_sc	0.16
406	elm_hberry	3.1
408	loak_ajun_pa	1.42
409	loak_ajun_wo	0.26
411	loak_mq_ajun_pa	2.62
413	mesq_loteb_sh	3.22
414	Mesquite_brush	0.22
415	oak_mq_jun_pawo	8.14
416	Other	7.78
418	pine_hardwood	2.35
419	poak_pawo	2.39
420	Poak_wo_for_gr	18.26
421	Post_oak_wofor	7.27
422	Silver_wintergr	5.32
423	Urban	2.33

424	Wiloak_waoak_gu	0.24
425	Woak_elm_hberry	1.39
500	Water	2.43

APPENDIX B

B.1 Projections:

lamutm15.prj (used in Arc/Info to convert a Lambert Azimuthal Projection to a UTM zone 15 projection)

Input
projection lambert
units meters
spheroid GRS1980
Parameters
34 55 0.000
27 25 0.000
-100 00 0.000
31 10 0.000
1000000.00000
1000000.00000
Output
projection utm
units 0.00100
zone 15
spheroid GRS1980
Parameters
End

albutm15.prj (Used in Arc/Info to project an Albers Equal Area projection into a UTM zone 15 projection)

Input
projection albers
units meters
spheroid clarke1866
Parameters
29 30 0.000
45 30 0.000
-96 0 0.000
23 00 0.000
0.00000
0.00000
output
projection utm
units 0.00100
zone 15
spheroid GRS1980
Parameters
end

B.2 Script

MergeTheme Script

Script was sometimes used to combine two coverages into one in Arcview. (Useful for combining the crops and forested areas. Crop land use codes were manually entered into new attribute table.)

’Name: View.MergeThemes

’Title: Merges two feature themes

’Topics: GeoData

’Description: Merges the selected themes into a single theme. A new

’shapefile is created which combines the shapes and attributes of the

```
'active themes.  The themes to be merged should have the same set of
'attributes (fields).  Only the fields from the first active theme are
'preserved in the output theme.
'Requires:  At least two themes of the same feature type must be in the
'active view.
'Self:
'Returns:
theView   = av.GetActiveDoc
theThemes = theView.GetThemes
if (theThemes.Count < 2) then
  MsgBox.Error( "Must have at least two themes in a view to merge.","")
  exit
end
'Allow the user to choose themes from the view to be merged...
themesToMerge = List.Make
while (true)
  t = MsgBox.Choice( theThemes, "Choose themes in view to merge:"+NL+
    "(Click Cancell to end):", "Merge Themes" )
  if (t <> Nil) then
    themesToMerge.Add(t)
  else
    break
  end
end
if ((themesToMerge = Nil) or (themesToMerge.Count < 2)) then
  MsgBox.Error("Not enough themes to merge.", "")
  exit
end
```

```
'Themes must have matching shape types for merging.  Using the first
'active theme verify that this is the case...
checkType  = themesToMerge.Get(0).GetFtab.FindField("Shape").GetType
for each i in 1 .. (themesToMerge.Count - 1)
  t = themesToMerge.Get(i)
  if (checkType <> t.GetFTab.FindField("Shape").GetType) then
    MsgBox.Error("Theme feature type mismatch - Unable to merge.","")
    exit
  end
end
'Specify the output shapefile...
outFName = av.GetProject.MakeFileName("theme", "shp")
outFName = FileDialog.Put(outFName, "*.shp", "Output Merged Shapefile")
if (outFName = Nil) then
  exit
end
'Create the list of fields used for the output theme.  The fields
'are taken from the first active theme only, it is assumed that
'other themes have an identical set of fields.  If this is not the
'case the themes will still be merged, however fields not found in
'other themes will be empty...
fieldList = List.Make
for each f in themesToMerge.Get(0).GetFTab.GetFields
  if (f.GetName = "Shape") then
    continue
  else
    fCopy = f.Clone
    fieldList.Add(fCopy)
```

```
  end
end
'Get the class of new FTab to create, create the new FTab and
'add fields that we've gathered from the input themes....
shapeType = themesToMerge.Get(0).GetFTab.FindField("Shape").GetType
if (shapeType = #FIELD_SHAPELINE) then
  outClass = POLYLINE
elseif (shapeType = #FIELD_SHAPEMULTIPOINT) then
  outClass = MULTIPOINT
elseif (shapeType = #FIELD_SHAPEPOINT) then
  outClass = POINT
elseif (shapeType = #FIELD_SHAPEPOLY) then
  outClass = POLYGON
else
  MsgBox.Error("Invalid shape  field type.", "Merge Themes"
  exit
end
mergeFTab = FTab.MakeNew( outFName, outClass )
if (fieldList.Count > 0) then
  mergeFTab.AddFields( fieldList )
end
'Populate the new FTab from the FTabs of the input themes...
for each t in themesToMerge
  av.ShowMsg( "Merging"++t.GetName )
  inFTab  = t.GetFTab
  if (inFTab.GetSelection.Count = 0) then
    theRecordsToMerge = inFTab
    numRecs = inFTab.GetNumRecords
```

```
  else
    theRecordsToMerge = inFTab.GetSelection
    numRecs = theRecordsToMerge.Count
  end
  for each rec in theRecordsToMerge
    av.SetStatus( (rec / numRecs) * 100 )
    newRec   = mergeFTab.AddRecord
    inField  = inFTab.FindField( "Shape" )
    outField = mergeFTab.FindField( "Shape" )
    mergeFTab.SetValue( outField, newrec, inFTab.ReturnValue( inField, rec ))
    if (fieldList.Count > 0) then
      for each f in fieldList
        fName    = f.GetName
        inField  = inFTab.FindField( fName )
        'Skip field if not found in inFTab...
        if ( inField <> Nil ) then
          outField = mergeFTab.FindField( fName )
          aValue    = inFTab.ReturnValue( inField, rec )
          mergeFTab.SetValue( outField, newRec, aValue )
        end
      end   'for each f
    end   'if count
  end   'for each rec
end   'for each t
av.ClearMsg
av.ClearStatus
if (MsgBox.YesNo("Add shapefile as theme to a view?",
  "Merge Themes", true).Not) then
```

```
  exit
end
'Create a list of views and allow the user to choose which view to
'add the new theme to...
viewList = {}
for each d in av.GetProject.GetDocs
  if (d.Is(View)) then
    viewList.Add( d )
  end
end
'Include a choice for a new view...
viewList.Add("")
addToView = MsgBox.ListAsString( viewList,"Add Theme to:", "Merge
Themes" )
'Get the specified view, make the theme, and add it...
if (addToView <> nil) then
  if (addToView = "") then
    addToView = View.Make
    addToView.GetWin.Open
  end
  mergeTheme = FTheme.Make( mergeFTab )
  addToView.AddTheme( mergeTheme )
  'Bring the View to the front...
  addToView.GetWin.Activate
end
```

BIBLIOGRAPHY

Arey, J., Crowley, D., Crowley, M., E., Resketo, M., and Lester, J. (1995) Hydrocarbon Emissions from Natural Vegetation in California's South Coast Air Basin, *Atmospheric Environment* Vol. 29, pp. 2977-2988.

Benjamin, M. T., Sudol, M., Bloch, L., and Winer, A. (1996) Low-emitting Urban Forests: A Taxonomic Methodology for Assigning Isoprene and Monoterpene Emission Rates, *Atmospheric Environment* Vol. 30, No. 9, pp. 2437-1452.

Benjamin, M. T., Sudol, M., Vorsatz, D., and Winer, A., (1997) A Spatially and Temporally Resolved Biogenic Hydrocarbon Emissions Inventory for the California south Coast Air Basin, *Atmospheric Environment*, pp. 3087- 3100.

Benjamin, M. T., and Winer, A. (1998) Estimating the Ozone-Forming Potential of Urban Trees and Shrubs, *Atmospheric Environment*, Vol. 32, No. 1, pp. 53-68.

Estes, M., Jacob, D., Jarvie, J., MacKay, J., and Smith, J. (1995) Biogenic Emissions Modeling for Southeastern Texas. Proceedings from Emissions Inventory Specialty Conference. New Orleans, LA.

Fuentes, J. D., Wang, D., Neumann, H. H., Gillespie, T. J., Hartog, G. D., and Dann, T. F., (1996) Ambient Biogenic Hydrocarbons and Isoprene Emissions from a Mixed Deciduous Forest, *Journal of Atmospheric Chemistry* 25: 67-95.

Geron, C. D., Guenther, A., and Pierce, T. (1994) An Improved Model for Estimating Emissions of Volatile Organic Compounds from Forests in the Eastern

United States, *Journal of Geophysical Research*, Vol. 99, No. D6, pp. 12773-12791.

Geron, C.D., Pierce, T. E., and Guenther, A. (1995) Reassessment of Biogenic Volatile Organic Compound Emissions in the Atlanta Area, *Atmospheric Environment*, Vol. 29, No. 13, pp. 1569-1578.

Guenther, A. B., Zimmerman, P., Harley, P., Monson, R., and Fall, R. (1993) Isoprene and Monoterpene Emission Rate Variability: Model Evaluation and Sensitivity Analysis, *Journal of Geophysical Research*, Vol. 98, No. D7, pp. 12609- 12617.

Guenther, A., Zimmerman, P., and Wildermuth, M. (1994) Natural Volatile Organic Compound Emission Rate Estimates For U.S. Woodland Landscapes, *Atmospheric Environment*, Vol. 28, No. 6, 1197-1210.

Lamb, B., Gay, D., and Westberg, H. (1993) A Biogenic Hydrocarbon Emission Inventory for the U.S.A. Using a Simple Forest Canopy Model, *Atmospheric Environment*, Vol. 27A, No. 11, pp 1673-1690.

Kinnee, E., Geron, C., and Pierce, T. (1997) United States land Use Inventory for Estimating Biogenic Ozone Precursor Emissions, *Ecological Applications*, 7(1), pp 46-58.

Monson, R. K., Lerdau, M., Sharkey, T., Schimel, D., and Fall, R. (1995) Biological Aspects of Constructing Volatile Organic Compound Emission Inventories, *Atmospheric Environment*, Vol. 29, No. 21, pp. 2989-3002.

Texas Parks and Wildlife Department, 'The Vegetation Types of Texas,' http://www.tpwd.state.tx.us/admin/veg/intro.html, Aug. 1996

Tingey, D. T., manning, M., Grothaus, L. C., and Burns, W. F. (1980) Influence of Light and Temperature on Monoterpene emission rates from Slash Pine, plant Physiology, Vol. 65, pp. 797-801.

U.S. Environmental Protection Agency, (1997) Urban Airshed Model (UAM) Biogenic Emission Inventory System Version 2 (BEIS-2) User's Guide, Air Quality Modeling Group, MD-14, Research Triangle Park, North Carolina, 27711.

Winer, A. M., Arey, J., Atkinson, R., Ashmann, S. M., Long, W. D., Morrison, C. L., and Olsyzk, D. M. (1992) Emission Rates of Organics from Vegetation in California's Central Valley, *Atmospheric Environment*, 26, pp. 2647-2659.

VITA

Irvin Wade Strange was born on August 13, 1973 in Silver City, New Mexico, the son of Nancy Ann Strange and Glenn Irvin Strange. He was an Eagle Scout and graduated co-valedictorian from Silver High School in May of 1992. In May of 1996 he received his Bachelor of Science degree in Mechanical Engineering from Southern Methodist University in Dallas, Texas. In August of 1996 he entered The Graduate School at The University of Texas at Austin to pursue a Master of Science degree in Environmental Health Engineering.

Permanent Address: 31 Rocky Creek Rd.
Silver City, NM 88061

This thesis was typed by the author.